RESEARCH & DEVELOPMENT

Roadway Signing and Marking of Unconventional Grade Separated Intersection Designs

Institute for Transportation Research & Education,
North Carolina State University

Christopher M. Cunningham
R. Thomas Chase
Kihyun Pyo

Department of Industrial & Systems Engineering,
University of Florida

David Kaber
Yunmei Liu

NCDOT Project 2019-26
FHWA/NC/2019-26
August 2022

Technical Report Documentation Page

1. Report No. FHWA/NC/2019-26	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Roadway Signing and Marking of Unconventional Grade Separated Intersection Designs	5. Report Date August 2022
	6. Performing Organization Code
7. Author(s) Christopher M. Cunningham, R. Thomas Chase, Kihyun Pyo, David Kaber, Yunmei Liu	8. Performing Organization Report No.
9. Performing Organization Name and Address Institute for Transportation Research and Education North Carolina State University Centennial Campus Box 8601 Raleigh, NC	10. Work Unit No.
	11. Contract or Grant No. NCDOT Project 2019-26
12. Sponsoring Agency Name and Address North Carolina Department of Transportation Research and Analysis Group 104 Fayetteville Street Raleigh, North Carolina 27601	13. Type of Report and Period Covered Final Report August 2018-August 2022
	14. Sponsoring Agency Code RP2019-26

15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. Abstract

This project developed driving simulator models and an experimental design to test signage placement impacts on driver capability to navigate three intersection types: Standard, Grade Separated Quadrant, and Grade Separated Contraflow. Driver performance was monitored using situational awareness and workload measures as well as operational measures such as speed deviation, lane changing, and deceleration.

Based on the research effort findings, it is recommended that signing engineers develop novel junction sign configurations, or provide additional guidance signs upstream of the decision point for intersections with non-traditional movements in order to offset low driver situational awareness and high cognitive workload, and to support timely lane changing behavior. The quadrant grade separated intersection design appears to be a feasible alternative to standard intersections with or without lane assignment signs and when using side-mounted decision point signs and providing lane information on the junction sign. Consequently, the results of this study provide some guidance for highway systems engineers on the need for novel signage designs to ensure effective driver information processing under unique highway configurations with performance comparable to standard intersections. From the data collected in this research, it is inferred that driver performance compared to standard intersections is similar at intersection forms with a non-intuitive turning movement (e.g. turn left to go right), whereas drivers at intersection forms which require advanced lane changing (e.g. contraflow and displaced left turns) may require additional guidance beyond that provided at standard intersections.

17. Key Words Grade Separated Intersection, Alternative Intersection, Driver Behavior, Signage and Marking, Human Factor, Driving Simulator	18. Distribution Statement No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 45	22. Price
Form DOT F 1700.7 (8-72)	Reproduction of completed page authorized		

DISCLAIMER

The contents of this report reflect the views of the author(s) and not necessarily the views of the University. The author(s) are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of either the North Carolina Department of Transportation or the Federal Highway Administration at the time of publication. This report does not constitute a standard, specification, or regulation.

ACKNOWLEDGEMENT

The research team acknowledges the North Carolina Department of Transportation for supporting and funding this project. We extend our thanks to the project Steering and Implementation Committee: Joe Hummer (Chair), Renee Roach (Co-Chair), Mustan Kadibhai (Research Engineer), Ayman Alquadwah, Clarence Bunting, John Sandor, and Katherine Hite. In addition, our team would like to acknowledge Shuyan Xia's work as an undergraduate student intern at the University of Florida. Her assistance with analytical and data collection tasks was very helpful to the team.

EXECUTIVE SUMMARY

Many grade separated intersection (GSI) forms are in service today across North Carolina. Traditionally, in busy urban and suburban environments with heavy traffic demands on mainline and side street arterial movements, grade separation has taken the form of partial and full cloverleaf designs. This often results in unnecessarily expensive infrastructure being constructed in an arterial environment not meant to provide the free-flowing movements you would expect to find in a freeway environment. Current ongoing research by this team is investigating a host of options that could be considered as an alternative to these more traditional designs – designs that very often require a smaller footprint, are likely to be cheaper to construct, and fit within the context of the intersecting roadways. It is anticipated that the results of this ongoing research effort will yield a handful of promising design alternatives the profession can consider under varying traffic demands and right-of-way constraints in a more urbanized environment where two arterials cross but need grade separation.

The first two research questions were partially supported by the findings:

(1) *Does the presence or absence of the lane assignment sign at a GSI influence driver visual behavior and situational awareness (SA) such that wrong way driving (WWD) frequency might be reduced?*
(2) *Does the position of a decision point sign at a GSI influence driver visual behavior and SA such that WWD frequency might be reduced?*

The use and placement of signs at the simulated GSIs did not result in significant differences in driver subjective responses such as SA and cognitive workload. However, the objective measures of driver performance and visual behavior were sensitive to the signage manipulations. To be more specific, use of lane assignment signs at interchanges did not appear to significantly increase driver SA nor decrease driver cognitive workload. Furthermore, overhead mounted decision point signs did not significantly increase driver SA or decrease cognitive workload, as compared to side-mounted signs. However, driver performance measures revealed that for the standard intersection and contraflow GSI, drivers primarily changed lanes after lane assignment sign use (Segment 2) or decision point sign (Segment 1) presentation. For the quadrant GSI, drivers primarily executed lane changes at the junction sign location (Segment 3) as quadrant junction signs also displayed lane information. Furthermore, analysis of the eye-tracking measures revealed percent change in pupil size (PCPS) to significantly increase when lane assignment signs were present and, consequently, this manipulation was more important for the contraflow design. The lack of significance of the decision point sign on driver behaviors could have been due to the overhead mount not being as visually accessible as expected, relative to the side-mounted signs.

Results partially supported the third research question:

(3) *Are there differences in driver visual and SA among various types of GSIs such that WWD frequency might vary among specific interchange designs?*

The contraflow design led to significantly degraded driver SA, likely due to a lack of driver familiarity with the upstream lane change configuration, as compared to the standard and quadrant interchanges. However, the quadrant design did not differ from the standard intersection in terms of SA. Results on cognitive workload revealed significant increases for drivers at both the contraflow and quadrant interchanges, as compared to the standard intersection. However, there was no difference between the contraflow and quadrant designs in terms of workload. Once again, these findings are likely due to the novelty of the GSI interchanges, lack of driver familiarity, and perceived complexity of navigation of the interchanges. In addition, driver performance responses showed the lowest Maximum Deceleration (MD) for the contraflow

design but the highest MD for the quadrant. The unique design of each scenario could have led to these results. Specifically, the deceleration segment of the contraflow design begins far from the point of interchange, while the quadrant and standard intersection designs have deceleration segments located at the periphery of the point of interchange. Related to this, the contraflow design resulted in many more drivers executing a lane change maneuver at or after the opening of the deceleration segment. We also found this interchange design to be highly correlated with the incidence of WWD through the machine learning model.

Based on the different observations of data, the location of the deceleration segment locations, and the additional lane information at the junction point sign for the quadrant as compared to the contraflow GSI, it is recommended that signing engineers develop novel junction sign configurations, or provide additional guidance signs upstream of the decision point for intersections with non-traditional movements, in order to offset low driver SA and high cognitive workload, and to support timely lane changing behavior. The quadrant GSI design appears to be a feasible alternative to standard intersections with or without lane assignment signs and when using side-mounted decision point signs and providing lane information on the junction sign. Consequently, the results of this study provide some guidance for highway systems engineers on the need for novel signage designs to ensure effective driver information processing under unique highway configurations with performance comparable to standard intersections. It is inferred driver performance compared to standard intersections is similar at intersection forms with a non-intuitive turning movement (e.g. turn left to go right), whereas drivers at intersection forms which require advanced lane changing (e.g. contraflow and displaced left turns) may require additional guidance beyond that provided at standard intersections.

Table of Contents

Tables

Figures

1. Introduction and Background

Rapid growth in transportation demands and limited funding for new roadways have resulted in serious roadway congestion issues, especially at urban intersections. Compounding this situation, there are locations where increasing infrastructure capacity using traditional means is not feasible or cost-effective. For example, according to a study by Eyler (2005), when the hourly traffic volume of an intersection exceeds 5,000 vehicles, the intersection needs to be expanded to three through lanes and double left-turn lanes. However, this way of thinking only enlarges intersections and can create complexity for signal operations and pedestrians, usually while also significantly increasing travel time variability and crash rates (Eyler, 2005). Considering the limitations of enlarging intersections and the need for reducing congestion and limiting roadway user conflicts, some researchers have considered other methods such as redesigning the configuration of intersections.

As a possible solution, departments of transportation have changed physical configurations of intersections (Leisch, 1993). These configurations, often referred to as alternative intersection and interchange (AII) designs, are intended to eliminate major conflict points by re-routing some movements to non-traditional patterns. Typical AII designs include roundabouts, reduced conflict intersections (RCI), continuous flow intersections (CFI), quadrant roadway intersections (QRI), and diverging diamond interchanges (DDI) (Brown et al., 2020). Among AIIs considered, some designs eliminate intersecting movements through grade separation. Grade separation provides vertical separation via one or more bridges and ramps resulting in increased capacity of the intersection. When the grade separation results in one or more uninterrupted roadways, it is referred to as an interchange. If both roadways remain signalized, it is a grade separated intersection (GSI). Leisch (1993) observed that GSI forms can generally be categorized into three types: diamond forms, cloverleaf forms, and rotary forms. The present research focuses on a driver simulation study of signage for the diamond and cloverleaf forms. We have previously addressed roundabout treatments in another simulation study (Salamati et al., 2012); therefore, rotary forms were not studied again in this effort. Among the diamond and cloverleaf forms, there are several subdivisions or varieties which are included.

1.1. Types of Interchange Designs

Since GSIs require grade separation, it is appropriate to review the impact of different types of interchange configurations on visual acuity. The primary motivation for this work is the observation that different types of interchanges produce different driver visual behaviors and vehicle control performance due to variations in navigation clarity (Morena & Leix, 2012). The most common and simplest form of interchange in the U.S. is the diamond interchange, which accounts for about 79% of the total number of interchanges (Atiquzzaman & Zhou, 2018). As each driving direction of the freeway or highway includes an off- and an on-ramp, the diamond interchange often has sufficient capacity for demands beyond what many at-grade intersections can serve. Diamond interchanges have multiple forms including single-point ("diamond" or "urban") interchanges (SPIs) and diverging diamond interchanges (DDIs), among others. However, a recent survey of the Departments of Transportation (DOTs) revealed that, due to construction and maintenance issues as well as the performance of alternative designs, it is unlikely that SPI designs will be considered as future alternatives unless the surrounding right-of-way and queue storage needed to adjacent signals warrants them (Brown et al., 2020).

Lloyd (2016) pointed out the DDIs can increase traffic capacity and reduce congestion by eliminating the need for left-turn phases. What's more, a recent evaluation of public perception of Missouri's first DDI revealed 80% of respondents believed the design to improve operations and safety (Chilukuri et al., 2011). In this evaluation, drivers also commented that they understood how the interchange worked. Claros et al.

(2015), using data from six DDI locations in Missouri, quantified the frequency of crashes across the entire DDI design. They found that DDI designs that replaced conventional diamond interchanges significantly decreased ramp terminal-related crashes. In addition, Claros et al. (2016) presented a case study revealing ramp terminals in a DDI design provide for a greater reduction in crashes than for the entire interchange footprint. Beyond this, Claros et al. (2017) also evaluated DDI ramp terminal safety by applying location-specific crash models. They compared fatal and injury, property damage, and total crashes. The authors concluded that DDI ramp terminals were safer than the conventional diamond signalized terminals. However, other issues have been identified for the DDI configuration. Cunningham et al. (2016b) stated that DDIs near adjacent signals could easily cause downstream spillback issues if not addressed holistically.

Many previous studies (Copelan 1989; Moler 2002; Braam 2006; Neuman et al. 2008; Morena & Leix, 2012; Zhou et al., 2012b; Vaughan et al., 2015; Atiquzzaman & Zhou, 2018; Abdelrahman, 2020) have shown that certain types of interchanges (such as diamond, cloverleaf, and partial cloverleaf) are problematic and susceptible to driver issues, including wrong-way driving (WWD). For instance, some studies observed that full diamond interchanges minimize driver confusion and mistakes in route selection. However, in some cases, drivers have been noted to confuse off-ramps of a diamond interchange with frontage roads running parallel to the ramp or highway, mistakenly turning the wrong way onto an off-ramp (Moler 2002; Braam 2006). Based on 6 years of crash data, Zhou et al. (2012b) identified that compressed diamond and diamond interchanges were the top two interchange types for WWD crashes. Related to this, Abdelrahman (2020) pointed out that one of the common concerns of DDI is WWD. Vaughan et al. (2015) used video cameras to monitor five DDIs and found that WWD tended to occur more often at the inbound crossover of the DDI.

Regarding the cloverleaf interchange configuration, this type of interchange can be further subdivided into full cloverleaf and partial cloverleaf. Although such designs are not as common in the U.S. as diamond interchanges, cloverleaf interchanges accommodate a large amount of traffic by supplementing collector and distributor roads in one direction or multiple directions of a freeway. Similar to diamond interchanges, WWD is also one of the main problems associated with cloverleaf interchange installations. Morena & Leix (2012) pointed out that a pair of on- or off-ramps in a partial cloverleaf, which are adjacent and parallel to each other, typically intersect with the crossroad at or near a 90-degree angle, making this configuration suspect for increasing the potential for driver confusion. For this reason, the partial cloverleaf design is regarded by many as more susceptible to WWD compared with diamond interchanges (Atiquzzaman & Zhou, 2018). Related, Baratian-Ghorghi et al. (2015) developed a mathematical method to estimate the likelihood of WWD events at exit ramp terminals for this type of interchange. They found that when left-turn volume turning onto the entrance ramp increased, and stopped vehicles departing the exit ramp decreased, the probability for WWD events increased. In addition, results also indicated that if a driver is more familiar with the interchange facility, the probability of WWD events also decreases.

Although WWD-related crashes are in the minority compared to other types of crashes at interchanges, they are associated with a higher probability of severe injuries and fatalities (Zhou et al., 2012a). Not surprisingly, there are many studies (Topolšek, 2007; Zhou et al., 2012a; Atiquzzaman & Zhou, 2018) that have conducted in-depth investigations of impacts of WWD on interchange safety. On this basis, WWD countermeasures have been proposed to reduce crash rates. However, the countermeasures differ in feasibility, applicability, effectiveness, implementation priority, and associated cost. Consequently, there is no one-size-fits-all solution to WWD events at interchanges. The NTTA (2009) has observed that some engineering measures could be taken to lessen the frequency of WWD events. For instance, Braam (2006) pointed out the common methods of manipulating roadway geometry and presenting signage and pavement markings. In addition, according to the results of National Cooperative Highway Research Program (NCHRP) Report 500 (Neuman et al., 2008), one of the prevalent strategies to minimize the possibility of WWD is to implement signage at freeway interchanges that are susceptible to WWD events. Moreover, Vaughan et al. (2015) also showed that key considerations in preventing WWDs in newly implemented DDIs include geometric manipulations and lighting and signage implementations.

1.2. Prior Research on Grade Separated Intersections

Grade Separated Intersections address the operational and safety gap between at-grade intersection designs and full interchange designs. NCDOT estimates at least 150 GSIs are already in existence in North Carolina with new designs under consideration. Limited guidance on GSI designs is available in FHWA's Alternative Intersections/Interchanges: Informational Report (Hughes et al., 2010) and two designs, the Echelon and Center Turn Overpass, are included in the Capacity Analysis for Planning of Junctions (Cap-X) Tool (Lochrane, Bared, and Zhang, 2011).

In recent research, Chase et al. (2020) identified 42 combinations of GSIs that are comprised of one of seven standard movement types on each of the crossing arterials. This research focused on the operations and safety of GSI designs, finding that many combinations performed better than at grade intersections across a number of turning movement volume trends. The report also identified the need to further research additional design considerations including signing and markings appropriate for these GSI designs. The movement types were direct left – downstream of bridge, direct left – upstream of bridge, single point left, three types of restricted crossing U-turn (RCUT with U-turn before Right turn, RCUT with Right turn before U-turn, and contraflow RCUT), and quadrant intersection. Out of these seven, the quadrant and contraflow designs were identified as particularly unique movements for drivers that may expect a traditional interchange design.

1.3. Prior Research to Address Problems

Although there have been a number of prior studies on interchanges, a select few investigations have focused on how to implement signage safely with different intersection configurations. Most studies address the design and evaluation of one or more specific interchanges. Leisch (1993) described several types of GSIs and discussed operational and design characteristics. Through a survey and interviews of drivers, Shumaker et al. (2008) observed that there are numerous documented unconventional intersection and interchange designs (e.g., GSIs). Shumaker et al. also noted that the public believes that these designs have the potential for driver confusion and few of them have been implemented. Shin et al. (2008) considered geometry, traffic demand levels, and signal conditions for four types of GSIs as controls for reducing congestion and promoting interchange safety. The GSI configurations included single-point interchanges (SPI), echelon interchanges (EI), center-turn overpasses (CTO), and two-level signalized intersections (TLSI). They concluded that the TLSI was an effective design to relieve traffic congestion. In a case study of the proposed Belgrade 'Hipodrom', Stanić et al. (2011) applied a three-step evaluation procedure, including a functional evaluation, micro-simulation, and multi-criteria evaluation, for two traffic solutions. The three solutions included complete grade separation of all intersection legs (the CPV alternative – 'grade-separated') and a grade separation designed to minimize construction costs (DMC 1 and 2 alternatives – 'minimize cost'). Results revealed the first functional evaluation step to show a small advantage for the DMC2 alternative, the micro-simulation indicated an advantage for the DMC1 alternative, while the multi-criteria evaluation was supportive of the CPV alternative vs. the DMC1. In addition, the DMC1 had the lowest construction cost.

Lloyd (2016) employed an Empirical Bayes (EB) analysis methodology in a study of collision rates for DDIs in Utah to characterize the impact of the interchange design on safety. Although the impact of the DDI designs on the frequency of collisions and safety outcomes was different for different interchanges, overall, the implementation of the DDI in Utah positively impacted crash occurrences. As can be noted from this review, most of the current research has focused on identifying factors contributing to WWD events at GSIs and developing predictive models. As additional examples, Wang (2018) quantified relationships between interchange geometric elements and WWD events at partial cloverleaf interchanges, and proposed several countermeasures for improving safety, such as stop line positions, median widths, etc. Furthermore, other research indicates that the association of infrastructure features with WWD events is magnified by specific driver characteristics, including driver's age, fatigue, distraction, and confusion (Copelan, 1989; Moler, 2002; NTTA, 2009; Zhou et al., 2012a,b). Last, other studies revealed that some countermeasures, such as lighting conditions and signage, have been insufficient and have actually contributed to WWD events (Braam, 2006; Vicedo, 2006). Unfortunately, there is lack of guidance in the literature on how to sign appropriately at various GSIs to achieve effective driver visual behavior and performance towards reducing incidents of WWD.

1.4. Absence of Signage Research

Regarding traffic signage research, there have been general studies of how to design and display signs at interchanges to support driving behavior. Richard & Lichty (2013) noted that when drivers are exposed to unfamiliar intersections and time-sensitive task demands, the need for clear navigation cues is even greater than at conventional interchanges. In other words, clear navigation signs should be developed to provide drivers with sufficient information for decision making before reaching a gore point. For freeway movements, appropriate signage and clear navigation information is even more critical due to vehicle speeds and the difficulty for drivers to correct mistakes if an exit is missed. In using such interchanges, drivers are making time-sensitive decisions and vehicle speed and traffic conditions can increase pressure and stress for drivers. Inman et al. (2006) evaluated the design of roundabout signs and found that as the number of items on signs increased, the accuracy of driver lane selection decreased significantly. However, their study was limited to the roundabout configuration and did not consider other types of intersections. Cottrell and Edara (2011) found that distance information on mainline specific service signs was useful to pilot study participants; however, their research focused on freeway approaches, not arterials. Vaughan et al. (2015) stated that DDI signage should be taken into careful consideration. They identified a need to strike a balance between "over-signing" and "under-signing" at interchanges. According to their results, under-signing can cause confusion due to a lack of driver direction and information and over-signing can cause confusion for drivers due to information load. Claros et al. (2016) conducted a site-specific safety assessment to quantitatively compare the safety of DDI ramp terminals with traditional diamond interchange ramps. They identified a number of important factors that may cause safety hazards, such as interchange geometry and the absence or presence of signs. In addition to presenting information on signs, the placement of signs can also affect driver behavior. Upchurch et al. (2005) pointed out that if drivers do not perceive a need to change lanes early enough, a lack of guidance signs can create performance problems.

The 2009 Manual on Uniform Traffic Control Devices (FHWA, 2009) provides standards, guidelines, and basic principles regarding how to install guide signs. However, the current manual does not provide guidance on how to install signs for different types of interchanges, which can greatly influence driver confusion and vehicle control behavior. Some prior pilot studies of sign and signal placement have been conducted. Zwahlen et al. (2003) conducted a field test of drivers negotiating six highway-to-highway interchanges. They found that with ground-mounted diagrammatic signs, test drivers who were not familiar with an area initiated lane changes to a correct lane earlier than without signage. At some interchanges, the lane change error improved by a factor of four to five times the before condition.

Similarly, Qiao et al. (2007) used driving simulation to study advance placement of guide signs for exits along highways. Based on physical sign location on the roadway and driver behaviors, the authors recommended an "optimal" advance placement of signs. Experimental results revealed that distance is crucial for installing guide signs, but Qiao et al. (2007) did not test whether ground-mounted signs and overhead signs have different effects on driver behavior. Using actual traffic data, Liao et al, (2014) proposed an algorithm to effectively and quantitatively evaluate the layout of interchanges and advance guide sign placement by calculating when the deviation rates from correct routes is smallest. Other recent studies have focused on the design of specific service signs on mainlines and ramps. Many of these studies (Carter, 2007; Zhang et al., 2013; Kaber et al., 2015; Zahabi et al., 2017) have compared different levels of information content and effects on driver distraction and performance under different driving conditions. However, none of these studies have investigated sign positioning or the content of more critical signs, such as lane assignment and decision point information, on ramps and at interchanges. In summary, how to properly display traffic signs at different GSIs in order to more effectively guide drivers without introducing distractions and WWD events has yet to be examined in the transportation literature.

1.5. Questions and Contributions

To address the above identified research gap on the design and placement of GSI signs, this study specifically examined the impact of traffic sign use and placement on driver visual behavior and situation awareness (SA) contributing to vehicle control performance. Driver vigilance decrements and degraded SA are considered to be precursors (proximate causes) of WWD occurrences. For this research, we elected to use a driving simulation method in order to analyze driver performance in an environment that is close to a naturalistic setting but that also allows for control of variables, such as vehicle traffic volume. We used a high-fidelity motion-base simulator to represent real-world driving conditions. The simulator also provided the advantages of low cost and low safety risk testing of multiple driving scenarios (as compared with observational studies). The specific research questions to be addressed by the simulator experiment were as follows:

(1) Does the presence or absence of lane assignment sign at a GSI influence driver visual behavior and SA such that WWD frequency might be reduced?
(2) Does the position of a decision point sign at a GSI influence driver visual behavior and SA such that WWD frequency might be reduced?
(3) Are there differences in driver visual and SA among various types of GSIs such that WWD frequency might vary among specific interchange designs?

In practice, it would be extremely challenging to test all kinds of GSIs; consequently, in this simulator study, we elected to expose participants to three types of intersections explained below and shown in Figure 1 (with Figure A showing the three sign types tested).

(1) At-grade Standard Intersection: The standard intersection in Figure 1A served as a control.
(2) Contraflow Intersection: The contraflow intersection represented movements where the left turn is presented upstream of the primary intersection. Signing tests for this intersection would also theoretically be applicable to the Continuous Flow Intersection, or CFIs (sometimes referred to as Divergent Left Turns, or DLTs). As shown in Figure 1B, drivers heading east must make an early left turn well upstream of the normal intersection location to continue heading north.
(3) Quadrant Intersection: The quadrant intersection uses one or more quadrants of the intersection to connect the two arterials. This is similar to the partial cloverleaf interchange options, but with signalization on both levels resulting in movements between levels possible on more quadrants. As show in Figure 1C, the research team and panel chose the opposing upstream quadrant for sign testing. In this way, drivers heading east must make an early left turn well upstream of the normal intersection, driving up or down to the crossing arterial, and turn left (in our case) to continue heading north.

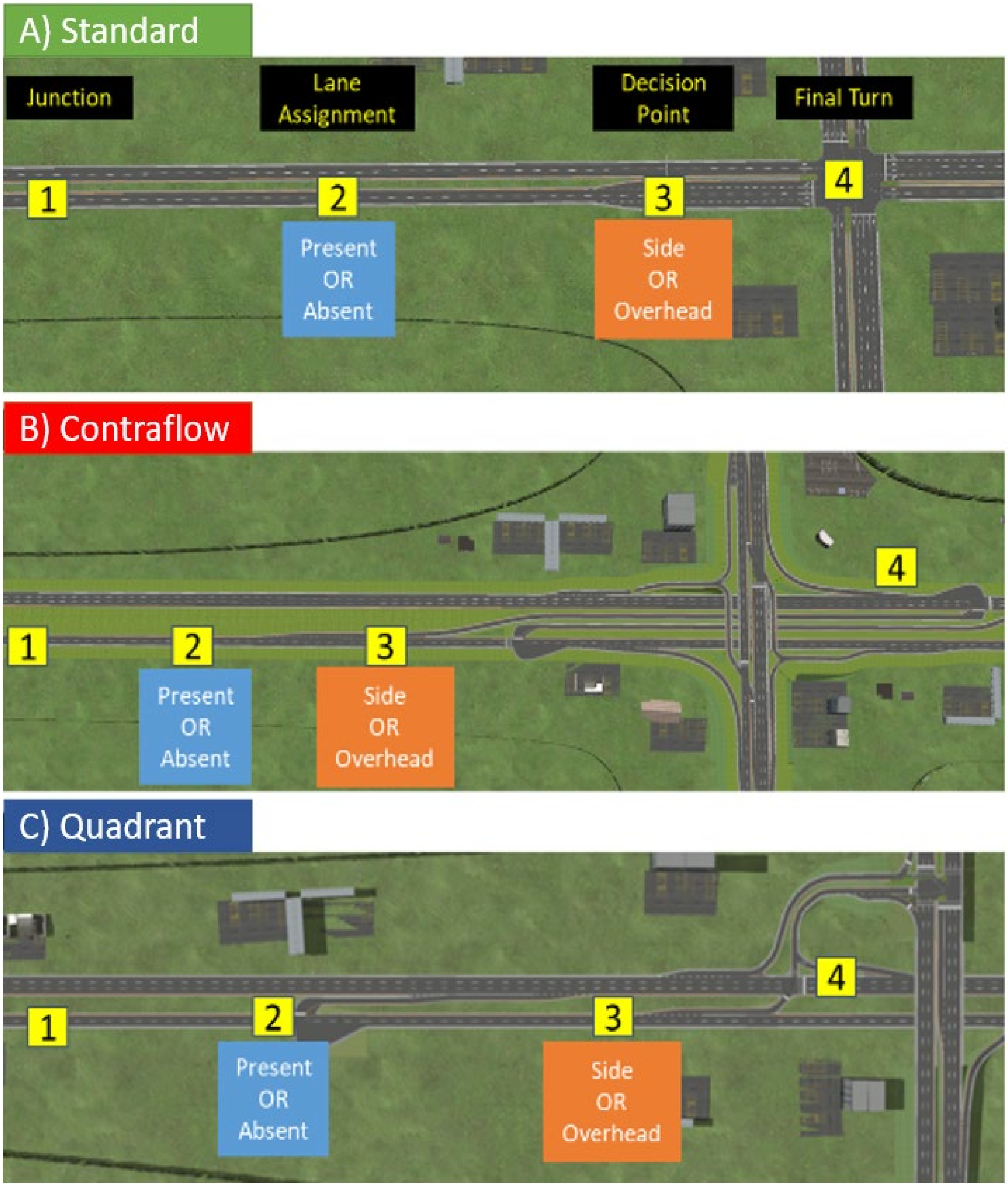


Figure 1: Three Grade-Separated Intersection Forms Studied

1.7. Hypotheses

Based on prior research on the use of different types of guide signs at conventional interchanges, three hypotheses were formulated.

(1) It was expected that the use of lane assignment signs would contribute to driver SA and driver performance.
(2) Considering driver visual attention patterns, it was expected that an overhead mount decision point sign would support greater SA and superior driver performance than a less visually accessible right-side mounted sign. An overhead mount sign aligns closer to driver foveal vision and right-side mounted signs fall in peripheral vision during normal driving making them less accessible.
(3) It was also expected that the contraflow and quadrant GSI designs would lead to degraded SA and driver performance, as compared to the standard intersection. This expectation is based on the complexity of the traffic flows and required gaze patterns in driver negotiation of contraflow and quadrant interchanges.

Beyond these expectations, we also anticipated that the various signage manipulations and interchange geometry variations would influence driver cognitive workload responses. Since cognitive workload is not a directly measurable human response, the most common approach to assessment is subjective survey of respondents, such as using validated task load indices. In addition, some other research has demonstrated various physiological responses to be useful for inferring cognitive workload states. The use of lane assignment signs in the driving scenarios was expected to moderate driver cognitive workload in negotiating the interchanges. Similarly, the visually accessible overhead mount decision point sign was expected to reduce cognitive workload responses. Lastly, the more familiar and less complex standard interchange design was expected to produce lower cognitive workload for drivers. Lower cognitive load was also expected to be associated with higher driver SA and improved vehicle control performance.

2. Method

2.1. Participants and Apparatus

A total of 48 participants with 20/20 vision (naturally or corrected) and a valid driver's license participated in the simulator experiment. The majority of drivers resided within a 25-mile radius of the North Carolina State University (NCSU) campus. All participants were compensated at a rate of $20 per hour. Participants were divided into two groups, according to age, with a convenience sample of young (18-24 years) and middle-aged (25-64 years) drivers. Due to the COVID-19 virus, under the guidance of the NC State Institutional Review Board (IRB), this experiment did not consider older (65 years and above) drivers. Gender was not applied as a grouping variable as similar recent studies (Deng et al., 2020; Zahabi et al., 2018) have demonstrated no effect of gender on driver behavior and performance in negotiating roadway circumstances with guide and specific service sign use.

A high-fidelity and full motion driving simulator (i.e., FORUM8 Co. Ltd, Tokyo, Japan) was used to investigate the effects of traffic sign usage and placement on driver behavior. During the simulator trials, drivers interacted with a full-size steering wheel, accelerator and brake pedal unit, a rear-view and two side mirrors, a dashboard, and in-vehicle traveler information message display. Drivers could control the simulated vehicle to maintain lane discipline, change lanes, and accelerate or decelerate. The dashboard and in-vehicle message display provided real-time feedback to drivers, including vehicle speed, engine RPMs (revolutions per minute), simulation time, vehicle position and direction information, and brake response. Surrounding the vehicle cockpit were eight 55-inch monitors providing drivers with a 360-degree road view to mimic the visual experience of a real-world driving environment. The vehicle cockpit and the visualization frame were mounted on a MOOG motion base, which allowed for a maximum payload of 900 kg (2,000 lb).

During test trials, participants donned a Pupil Labs eye tracking headset to record their eye movement activity. This particular tracker records a driver's pupil movements at a frequency of 200 Hz. The recording is combined with the driving scene view with a viewing angle of 100 degrees diagonal captured at 120 Hz.

2.2. Experiment Design

This study followed a 2×2×3×2 mixed within-subject and between-subjects experiment design, with two lane assignment sign settings (present and absent), two types of decision-point sign positions (side and overhead), three types of interchange designs (standard, contra-flow, and quadrant), and two driver age groups (young and middle-aged). Each participant was assigned one unique sign combination with exposure occurring across the three types of geometric layouts with replication. In total, there were six trials for each participant.

Crossing the two-lane assignment sign settings and the two decision point sign positions yielded four types of sign combinations (see Table 1). For example, one sign option assigned to a participant included the presence of a lane assignment sign with an overhead decision-point sign configuration. The participant was then repeatedly exposed to this particular combination across the various geometries. Exposure to the designs was replicated to assess within-subject performance variability. Consequently, each participant completed two at-grade standard intersection trials, two contra-flow GSI trials, and two quadrant GSI trials. In order to sequence the test trials and limit any carryover bias among trials and specific conditions, we used a Latin square randomization procedure. We selected two 3x3 Latin squares (from among 12 possible squares) to serve as the basis for unique scheduling of six test trials for a block of three subjects (within an age group and assigned to a specific signage combination). The first square was used to randomly assign the initial presentation of each interchange design for all participants. The second square was used to randomly assign the replicated exposure to each interchange condition. For the replication trials, the first test did not repeat the last test as part of the initial interchange exposure and no pairings of interchange designs appearing in the initial exposure ordering sequence were found in the sequence of replication trials. After obtaining the test trial sequences for three participants, we applied the same randomization procedure to define test trial orders for all remaining participants in each age group with assignment to a specific signage combination. This approach allowed for assessment of between-subject performance variability within an age group and for a target signing condition.

Table 1: Sign Options Configuration

4 Types of Sign options	2nd sign: Lane Assignment	3rd sign: Decision Point
Sign Option 1	Present	Side
Sign Option 2	Present	Overhead
Sign Option 3	Absent	Side
Sign Option 4	Absent	Overhead

2.3. Experiment Procedures

Prior to participating in the driving simulator experiment, participants completed an informed consent form, as required by the NC State IRB, a Pre-Trial Driver Background Questionnaire (PTDBQ), and a Pre-exposure Simulator Sickness Questionnaire (PSSQ). The questionnaires provided the research team with information on participant physical and mental health before the experiment, which served as a benchmark for sickness assessments during experiment trials. Participants were also required to present their valid driver's license before testing.

Subsequently, each participant completed simulator driving training, which was combined with calibration of the eye tracker. The driver training included three phases with right turns, proceeding straight on a highway, and left turns. Participants drove towards a T-intersection and turned right. They accelerated to 45 mph and maintained this speed on a straight section. They then reduced speed to safely navigate a S-curve section. They accelerated to 45 mph and maintained this speed until they reached an overpass, where they reduced speed and entered a left-turn bay to turn onto a highway ramp. Finally, they accelerated to a sufficient speed to merge with free-flowing highway traffic. Figure 2 shows the training scenario views.

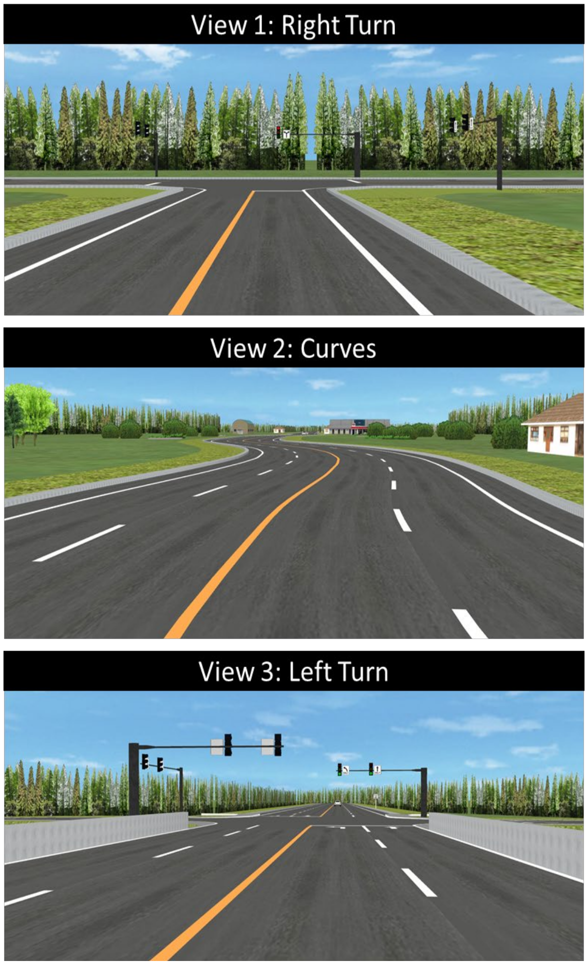


Figure 2: Training Scenario

With respect to use of the eye tracker, screen marker calibration was applied in this study. Participants gazed at markers on the simulator display screens to complete the calibration. After the driving training and calibration, participants completed a workload demand component ranking form as part of the NASA-Task Load index (NASA-TLX) methodology (please refer to 2.5.2 for details).

For experiment test trials, participants drove on a short segment of highway and were asked to stop at a specific location, including a stop sign or signalized intersection point. At these stopping points the driving simulation was frozen and participants were posed with a tablet-based workload analysis form and SA questionnaire. Participants immediately began rating driving workload demands. At the same time, an experimenter shut-down all the simulator display screens and drivers subsequently responded to the SA questions. The research team developed a data collection sheet for each specific experiment condition, which was used to record participant responses to the SA queries, as well as the ground truth of the driving simulation. (This information was used for data analysis purposes.) Once the surveys were complete, the simulator displays were reactivated and the driving simulation resumed until completion of the scenario (driver route selection based on signage).

Finally, there was a 10-minute break after every two test trials. During each break, participants completed a Simulator Sickness Questionnaire (SSQ) to ensure that they did not suffer any symptoms during the experiment. For each participant, the entire experiment took approximately 1.5 hours.

2.4. Driving Task

The driving simulator presented an urban environment and a medium-sized car (sedan) placed in the right-most of two lanes. Participants were given a destination of "Garden St, North" for all test trials. For each scenario, there were four possible locations in which signage could appear (Positions 1-4). Position 1 presented Junction Information and Position 4 presented Final Turn Information without change across all trials. However, Position 2 was used to present the lane assignment sign (if present) and Position 3 presented the decision point sign (overhead or right-side). Figure 3 shows the exact signage locations which were determined for each intersection geometry.

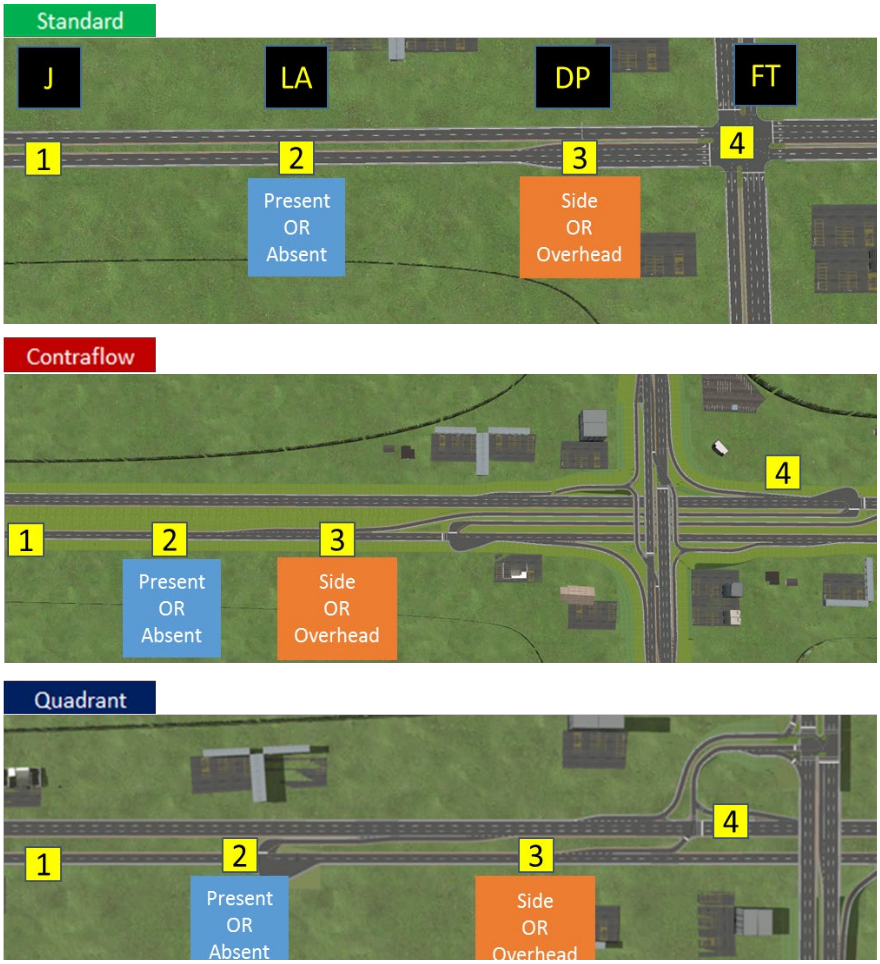


Figure 3: Three Intersection Configurations with Sign Locations

Consistent instructions were given to each participant driving through each of the 3 intersection forms tested. Figure 4 provides aerial and plan views for reference followed by an explanation of the instructions and (correct) movements that would be made if the participant navigated the intersection correctly.

Figure 4: Aerial and plan view of each scenario.

During each at-grade *standard intersection* test trial, drivers were to follow a two-lane roadway and maintain their vehicle speed at a limit of 45 mph. They were also directed to exhibit normal driving behavior, such as lane selection, until seeing further destination guidance information. Drivers were permitted to make lane changes based on ambient traffic and to decelerate to enter a left-turn lane at the intersection and proceeded to the signal. Following a full stop, drivers waited for traffic and/or a signal to turn left. After turning left onto a two-lane segment, drivers once again accelerated to 45 mph and maintained this speed until the end of the experiment scenario.

During each *contra-flow GSI* test trial, drivers were to follow a two-lane roadway and maintain their vehicle speed at a limit of 45 mph. They were also directed to exhibit normal driving behavior, such as lane selection, until seeing further destination guidance information. Drivers were permitted to make lane changes based on ambient traffic and to decelerate to enter a left-turn lane upstream of the intersection and proceeded to the signal. Following a full stop, drivers waited for traffic and/or a signal to make a U-turn onto a three-lane segment. Drivers once again accelerated to 45 mph and maintained this speed until they exited onto a right-side ramp per the signage. After exiting the ramp, drivers accelerated to 45 mph and maintained this speed until the end of the experiment scenario.

During each *quadrant GSI* test trial, drivers were to follow a two-lane roadway and maintain their vehicle speed at a limit of 45 mph. They were also directed to exhibit normal driving behavior, such as lane selection, until seeing further destination guidance information. Drivers were permitted to make lane changes based on ambient traffic and to decelerate to enter a left-turn lane at the intersection and proceeded to the signal. Following a full stop, drivers waited for traffic and/or a signal to turn left. After turning left onto a two-lane segment, drivers once again accelerated to 45 mph and maintained this speed until they entered a left-turn lane at the second intersection and proceeded to the signal. Following a full stop, drivers waited for the signal turn left onto a two-lane segment. After turning left, drivers accelerated to 45 mph and maintained this speed until the end of the experiment scenario.

2.5. Response Measures

As a basis for analyzing driver behavior in negotiating the various interchange configurations, we collected data on several different response measures during the experiment test trials. The measures included a SA assessment questionnaire, a cognitive workload demand survey, driver performance/vehicle control variables, and visual behavior measures.

2.5.1. Situation Awareness

Endsley (1995) defined SA as the perception of elements within a volume of time and space (Level 1), comprehension of their meaning (Level 2), and projection of their status in near future (Level 3). SA measurement methods vary among studies. The Situation Awareness Rating Technique (SART; Taylor, 1997) and Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995) have been frequently applied in research. SAGAT is considered to be an objective method of measuring SA, as it involves queries of operator dynamic knowledge in real-time. SAGAT is also considered to be a global measure of SA as queries are targeted at all three levels, as identified by Endsley. Therefore, SAGAT was applied in this study to measure driver SA at the various interchange configurations. According to Jones and Kaber (2004), one of the keys to successful SAGAT data collection is the efficacy of the design and development of a query database.

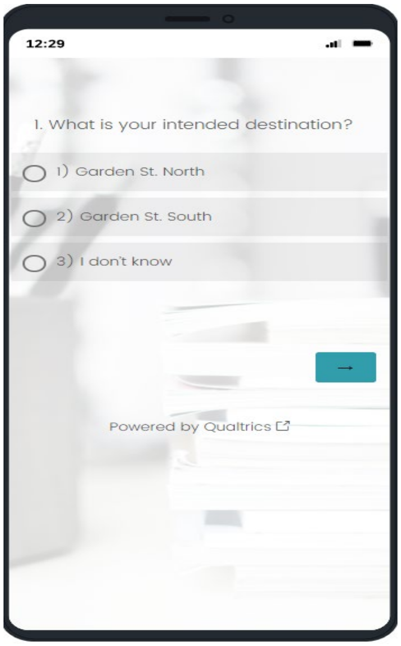
On this basis, we worked with a subject matter expert (SME) to brainstorm potential queries pertaining to all elements of the roadway environment and driving task, including signs, lanes, speeds, signals, vehicles, direction of travel, travel time, distances, and driver behaviors. In addition, we ensured that queries could be applied across driving scenarios. This process resulted in a pool of 29 queries for each simulation scenario (interchange configuration). While all queries were applicable across scenarios, their answers varied from scenario-to-scenario, based on the interchange design. Subsequently, human factors experts categorized the candidate queries according to three levels of SA, resulting in 13 queries for Level 1 and 8 queries for each of Level 2 and Level 3.

Next, we developed sets of possible answers to queries (29 queries * 3 scenarios = 87 possible answers). Most answers can be prepared in advance of experiment trials. For example, if a driver is asked how many green guide signs they saw before a stop (Level 1), the answer can be determined based on the simulation design. However, for some queries, driver answers may depend on simulation events. Consequently, there is a need for experimenters to record the state of the simulation during a test trial. For example, a driver may be asked what his or her average speed was before slowing down for the first turning motion (Level 2). All answers should be plausible and based on the simulation design or events. To moderate the effect of random guessing on statistical results, each query also included an "I don't know" option and research team members reminded subjects of the option before each test trial.

Finally, the terminology of all queries was reviewed for accuracy and clarity by an expert highway systems program manager and driving simulation programmer. We also conducted pilot tests of queries with 2-3 transportation researchers. Based on these tests, we resolved any ambiguities in SA queries and answers. Ultimately, we ensured that there was only one correct answer for each of the SA queries.

For presentation of the SA questionnaires during test trials, we used Qualtrics survey software. Queries were randomly selected from a large pool customized to the specific driving scenario. At each vehicle stop during a test trial, 6 queries (from among 29) were posed to drivers with two queries representing each level of SA. Based on our experiment design with 6 different scenarios, 6 trials per scenario and 6 queries per trial, there were a total of 216 observations on driver SA for each age by signage group. The Qualtrics surveys presented a single query per application screen to prevent subjects from obtaining information to answer one query based on another. We also required subjects to provide an answer to a query before advancing to the next question to ensure some datapoint was available for grading all queries. There was no “back button” to prevent participants from modifying answers to queries, after reading another query (see Figure 5for an example interface).

Driver responses to the SAGAT queries were graded based on the ground-truth of the driving simulation. For each query, there were only two scoring outcomes: 0 if not answered correctly or the option of “I don't know” was selected; 1 if answered correctly. On this basis, the percentage of correct responses for each trial was calculated as a SA score.



1. What is your intended destination?

○ 1) Garden St. North

○ 2) Garden St. South

○ 3) I don't know

Figure 5: Qualtrics Survey software allowed for questionnaire to be ported to tablet or cell phone app.

2.5.2. *Driver Workload*

According to previous human factors studies (Endsley & Kaber, 1999; Kaber & Endsley, 2004; Fuller, 2005; Zahabi et al., 2019), the NASA-Task Load indeX (TLX; Hart & Staveland, 1988), the National Aeronautics and Space Administration Task Load Index (NASA TLX) is a commonly used measure of cognitive workload and has demonstrated reliability. The purpose of using this index was to determine the cognitive load imposed on drivers by the signage conditions when negotiating the various types of interchanges. At the beginning of the experiment, participants ranked the importance of six workload demand components, including mental, physical, temporal, performance, effort and frustration, for the driving task. At the end of each test trial, participants rated their perceived mental workload, according to six demand components using a 100-point scale. The NASA TLX was calculated as the rank-weighted sum of the demand ratings scaled from 0 to 100 points.

2.5.3. *Driver Performance*

Before discussing subjects' driver performance, it is necessary to declare two distance concepts that will be addressed in this report: (1) Foveal distance is distance from the point where the driver can read the words on sign to sign; (2) Deceleration section is section from the beginning of taper to where the dotted lane line ends.

The driving simulator automatically collected vehicle speed and position information throughout each test trial, leading to three responses: absolute speed deviation (ASD) from speed limit, maximum deceleration (MD), and lane change position (LCP). The decision point signs were located in deceleration segments; consequently, it was not possible to assess the ASD response during decision point sign exposure. For the ASD response, we focused on driver exposure to junction (J) point and lane assignment signs when at foveal (or reading) distances from the signs. As for the LCP response, our overarching objective was to assess how far drivers were from the deceleration segment when changing lanes. Therefore, we counted the number of drivers occupying the left lane every 100 meters between the start of the experiment and the start of deceleration segments. The opening point of the left turn lane marked the initial point of interest and then we measured upstream from there until reaching the initial vehicle position for the experiment.

2.5.4. Percentage Change of Pupil Size

In addition to the NASA TLX assessment of driver workload, we also sought to capture a real-time objective indicator of cognitive load by using a physiological response. Driver pupil diameter and the percent change in pupil size (PCPS) have been widely used for insight into cognitive states, including arousal, attention demand, and cognitive workload (Zhang et al. 2016; Attard-Johnson, 2019; Kret & Sjak-Shie, 2019; Zahabi et al., 2019; Zahabi et al., 2021). Consequently, we measure the PCPS for the various types of sign combinations in each interchange configuration. Pupil position and size data were output from the Pupil Labs device and was post-processed to generate the PCPS response. Average values of PCPS were calculated for each driving condition by using a custom algorithm in Python.

2.6. Analysis and Models

2.6.1. Descriptive and Inferential Statistical Analyses

Descriptive statistical analyses were initially performed on all response measures, including calculation of means and standard deviations, and preparation of graphical analyses, such as histograms. These analyses characterized the distribution of data sets, identified whether there were missing values or outliers, provided insight into any relationship between the independent and dependent variables, and revealed any interactions among the independent variables.

For inferential statistical analyses, we applied various analysis methods based on the types of dependent variables. For continuous responses that upheld the parametric test assumptions (i.e., the NASA TLX, ASD, MD, and PCPS), we applied analysis of variance (ANOVA) along with Tukey's post-hoc tests. For other discrete quantitative variables or variables that did not support parametric test assumptions (i.e., SA, LCP), we leveraged non-parametric tests, including the Kruskal-Wallis rank sum test and Wilcoxon rank sum test.

2.6.2. Machine Learning Model Training and Evaluation

In addition to statistical inference, this study also developed machine learning models to analyze and predict driver behavior based on interchange configuration and signage combinations. Machine learning models have become popular for analyzing driver behavior. Zouhair et al. (2020) said that machine learning for driver behavior research typically falls into five categories, including: vehicle-based models, physiological-based models, behavior-based models, subjective models, and environmental measurement models. Vehicle-based models address driver ability to control a vehicle, including speed, acceleration, and lane changes. Physiological models involve recording driver physiological states, including muscular responses

(EMG), brain signal responses (EEG), cardiac responses (ECG), and skin conductance responses (GSR). Behavior-based models refer to capturing driver attitudes, facial expressions, and verbal responses via audio or video processing techniques. Subjective models are based on driver self-reports of status, including the NASA-TLX and SA responses. Finally, environmental-based models refer to capturing temperature and humidity, dry and wet road conditions, and the design of the road for predicting driver behaviors in response to critical driving events.

Regarding applications, driver physiological and subjective models can be used to predict stress states or specific driver emotions. Such models can also provide a basis for determining whether certain events will lead to operational errors, like wrong way driving. Osman et al (2019) tested and compared the accuracy of various machine learning models for identifying driver secondary task performance according to vehicle-based measures (speed, longitudinal acceleration, lateral acceleration, pedal position, and yaw rate). They found that a Random Forest (RF) technique produced the highest classification accuracy (82.2%). Yeo et al. (2009) applied a support vector machine (SVM) classifier to driver EEG signals and they were able to identify normal wakefulness vs. light drowsiness with over 99% accuracy. Jabon et al. (2010) applied multiple machine learning models to identify key facial features (from videos) at different pre-incident intervals. The feature extraction was then used to predict minor and major accidents. They found the logitBoost classifier to produce the highest performance accuracy for this purpose. Halim and Rehan (2020) applied three different classifiers (SVM, neural network (NN), and RF) to driver EEG signals for identifying driving-induced stress with experimental conditions and the EEG responses labeled based on self-reports of stress. Lastly, Zahabi et al. (2021) applied SVM, RF and Random Fourier Feature (RFF) machine learning classification algorithms to driver eye tracking and behavior data to classify states of police officer distraction during in-vehicle technology use.

Based on the Zouhair et al. (2020) research, for the present study we used vehicle-based measures (i.e., driver vehicle control performance) as a basis for labeling/classifying test trials as erroneous or error-free. We also use subjective and environmental-based measures as features to make prediction of driver errors in test trials.

Data Preparation. The purpose of our model was to predict whether the design of interchanges and the use and placement of road signs is associated with erroneous driving maneuvers, given different driver mental states (i.e., SA and mental workload). In this way, the machine learning models link underlying driver behaviors to WWD incidents. The input variables to the machine learning model were divided into four categories, including: driver characteristics and responses (age group - young or middle age, SA score, and mental workload); and roadway environment conditions (scenario types Standard, Contra-Flow, or Quadrant), lane assignment sign status (present or absent), and decision point sign position (side or overhead)).

Labeling. We classified driver performance in two categories: error or error-free. We further divided erroneous performance into three sub-categories based on ASD, MD, and LCP.

(1) If a driver's ASD was greater than or equal to 6 mph, we labeled his (or her) performance as erroneous. Most states in the U.S. do not impose fines for speeding less than 5 mph;
(2) If a driver's MD was more than 15 ft/s^2, we labeled his (or her) performance as erroneous. Wortman and Fox (1986) stated that MD (and maximum acceleration) should be less than 15 ft/s^2 to ensure driver safety;
(3) If a driver's LCP occurred during or after the deceleration segment, we labeled his (or her) performance as erroneous. We expected drivers to change lanes as directed by the road signs before the deceleration segment. It is worth pointing out that WWD was also a subset of the LCP error condition.

Training process. We made comparison of four classifiers commonly used in driver behavior research, including SVM, k-nearest neighbor (KNN), decision trees (DT), and RF. We randomly divided the entire experiment dataset into two parts for model training (80%) and testing (20%). In addition, with a very small sample dataset (288 data points), we applied a 10-fold cross-validation method. To avoid the influence of different feature value ranges on model weights assigned to training data patterns, we normalized all numerical input variables and transformed categorical data to binary responses with “one-hot encoding”.

Model evaluation. According to Zouhair et al. (2020), accuracy and sensitivity (also known as recall) are the most common evaluation methods considered in driving behavior research with 65.85% of studies using accuracy and approximately 35.36% using sensitivity. For binary classification, accuracy can be represented in terms of true/false positives (TP/FP) and true/false negatives (TN/FN) as in Equation (1). Sensitivity/recall can be calculated as in Equation (2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

In these equations, a TP represents the number of driving errors that were accurately predicted; a TN is the number of no driving errors that were accurately predicted; a FP (or false alarm) is the number of driving errors that were not present but predicted (i.e., a Type I error); and a FN is the number of driving errors that were not detected (i.e., a Type II error). The accuracy response for the model represents the probability of accurately predicting erroneous or error-free driver performance; while recall represents the probability that a driving error is accurately predicted if an error actually occurs.

In addition to these responses, the Area under Curve (AUC) is also a well-known measure for evaluation of machine learning models, especially binary classification models. It is an aggregate measure of the discriminative power of a predictive model across all possible classification thresholds (output values differentiating between erroneous and error-free performance). In terms of probability theory, AUC is the likelihood that a ML model ranks a random positive case more highly than a random negative example (Géron, 2019). The advantage of AUC is that it takes into account a classifier's capability to classify positive and negative cases. In the case of unbalanced samples, a classifier can still be reasonably evaluated using the AUC measure. The higher the AUC, the higher the discriminative power and the better the model. When AUC equals 0.5, the classifier is considered to be a random classifier providing no additional information for an analyst.

3. Results and Discussion

3.1. Driver Profile

In recruiting subjects for the experiment, 24-years of age was used as a boundary to distinguish between young and middle-aged drivers; that is, 18 to 24 years was considered young and 25 to 64 years was considered middle-aged. Pareto analysis (Figure 6) indicated that 33 subjects were between the ages of 18 and 26, accounting for 68.75% of all subjects. On the contrary, there were only 6 subjects over the age of 34, accounting for 12.5% of the sample. Consequently, there was a limited actual age between the groups, which may further explain the observed effects of age on all responses.

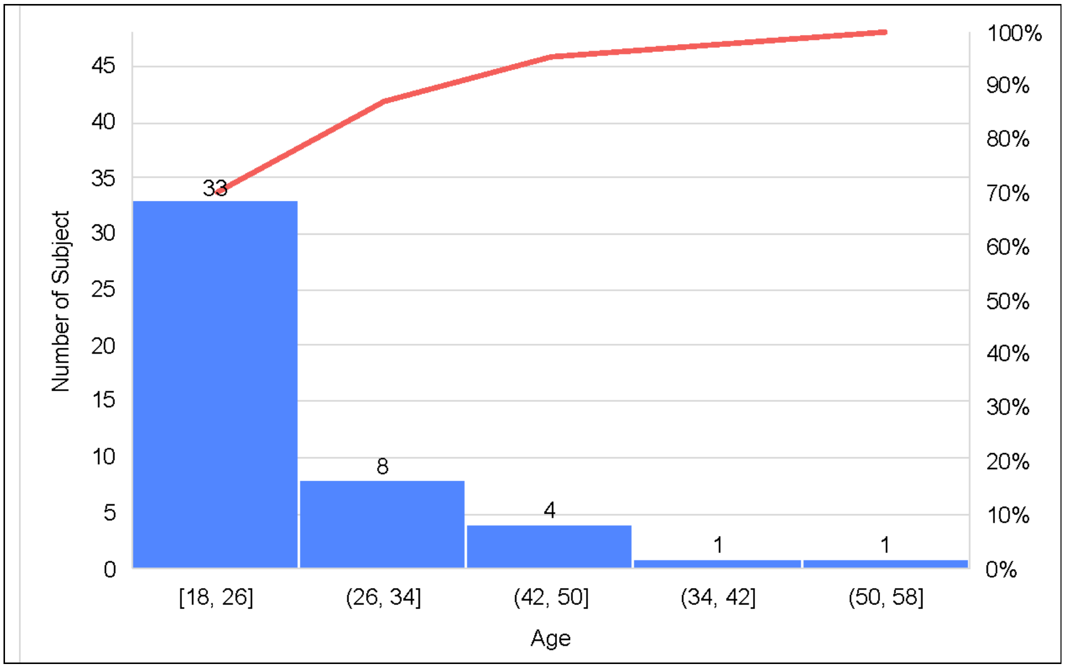


Figure 6: Pareto chart of subject's age distribution.

3.2. Situational Awareness

Across test trials, there were 1728 queries administered with 1114 correct responses and 614 incorrect responses. There were no missing values in the SA dataset. The percentage of correct responses was calculated for each trial. Among all participants, the average SA score was 0.65 (indicating greater than "chance" response of 0.5) and the standard deviation was 0.22.

Situational Awareness (SA) describes the driver's perception and comprehension of their surroundings as well as their prediction of surroundings they may encounter in the immediate future"

Figure 7 (a)(b)(c)(d) reveals the mean SA responses for the different driving scenarios, age groups, and the use and placement of signs, accordingly. Situation awareness differences among the scenarios were pronounced, including: Standard: 0.71; Quadrant: 0.67; and Contraflow: 0.57. Conversely, age group and the use and placement of signs appeared to have little effect on driver SA score. Since the SA score was composed of three levels, we also descriptively analyzed the relationship between the rate of wrong answers at each level of SA and the different independent factors (age groups, scenarios, and use and placement of signs).

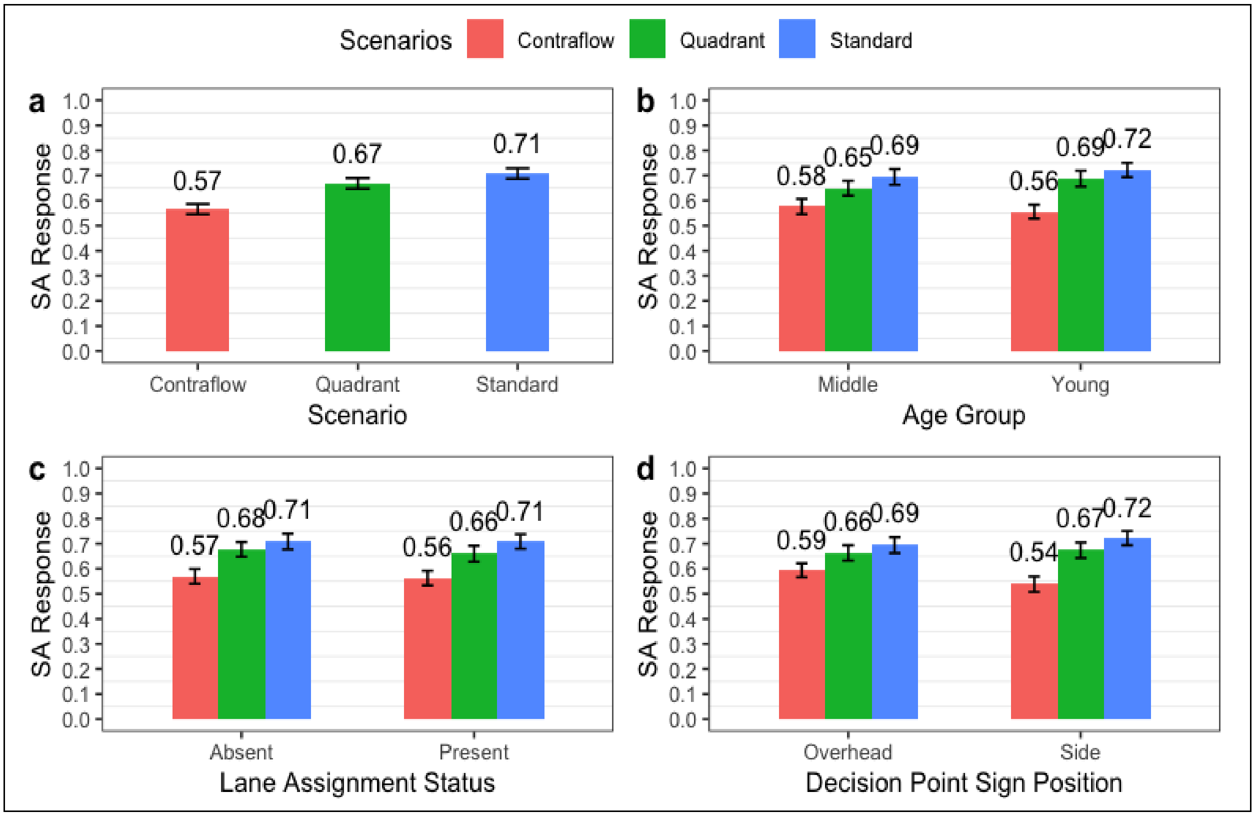


Figure 7: (a) Mean of SA response for different scenarios; (b) mean SA response by scenario type and age group; (c) mean SA response by scenario type and lane assignment sign use; (d) mean SA response by scenario type and decision point sign position.

As shown in Figure 8 (a)(b)(c)(d), the rate of wrong answers appeared to vary among levels of SA according to scenario type and the lane assignment sign condition. The Contraflow configuration appeared to yield consistently low driver SA across levels; whereas, the Standard and Quadrant interchanges appeared comparable for perception and roadway/vehicle state comprehension. The Standard intersection appeared superior for projection (route selection), likely because the standard intersection is the most frequently encountered intersection form. Regarding the lane assignment sign condition, Level 1 and 2 SA appeared to be comparable for sign and no sign settings. It is possible that the presence of a lane assignment sign may have supported drivers in Level 3 SA (projection of routes).

> The Contraflow intersection yielded the lowest *situational awareness* responses compared to the Quadrant and Standard Intersections.

To make statistically supported inferences on the SA scores for the test conditions, we applied a mixed-effects model to the data. Due to the limited number of SA queries per trial, the response was discrete. Consequently, data revealed a parametric test assumption (normality) violation, specifically a Q-Q plot with banding and a Shapiro–Wilk's test at $p < 0.01p < 0.01p < 0.01p < 0.01$. Therefore, we turned to non-parametric methods, including a Kruskal-Wallis rank sum test and Wilcoxon rank sum test to analyze the SA data.

The Kruskal-Wallis rank sum test revealed significant differences in driver SA under different interchange conditions ($\chi^2 = 22.12,\ p = 6.07 \times 10^{-4}$). We conducted pairwise condition comparisons using the Wilcoxon rank sum test. Results revealed driver SA in the Contraflow scenario to be significantly lower than for the Standard interchange ($W = 6312.5, p = 5.61 \times 10^{-6}$) and Quadrant ($W = 5830.5, p = 1.10 \times 10^{-3}$). There was no significant difference in driver SA between the Standard and Quadrant scenarios ($W = 5117.5, p = 0.17$). Furthermore, there were no significant differences in SA among age groups ($\chi^2 = 2.99,\ p = 0.81$) and the use of lane assignment ($\chi^2 = 3.21,\ p = 0.78$) and placement of decision point signs ($\chi^2 = 2.31,\ p = 0.89$). The non-parametric methods did not support analysis of interactions of intersection design settings and use and placement of signs in terms of driver SA, preventing analysis of the differences in placement at different designs.

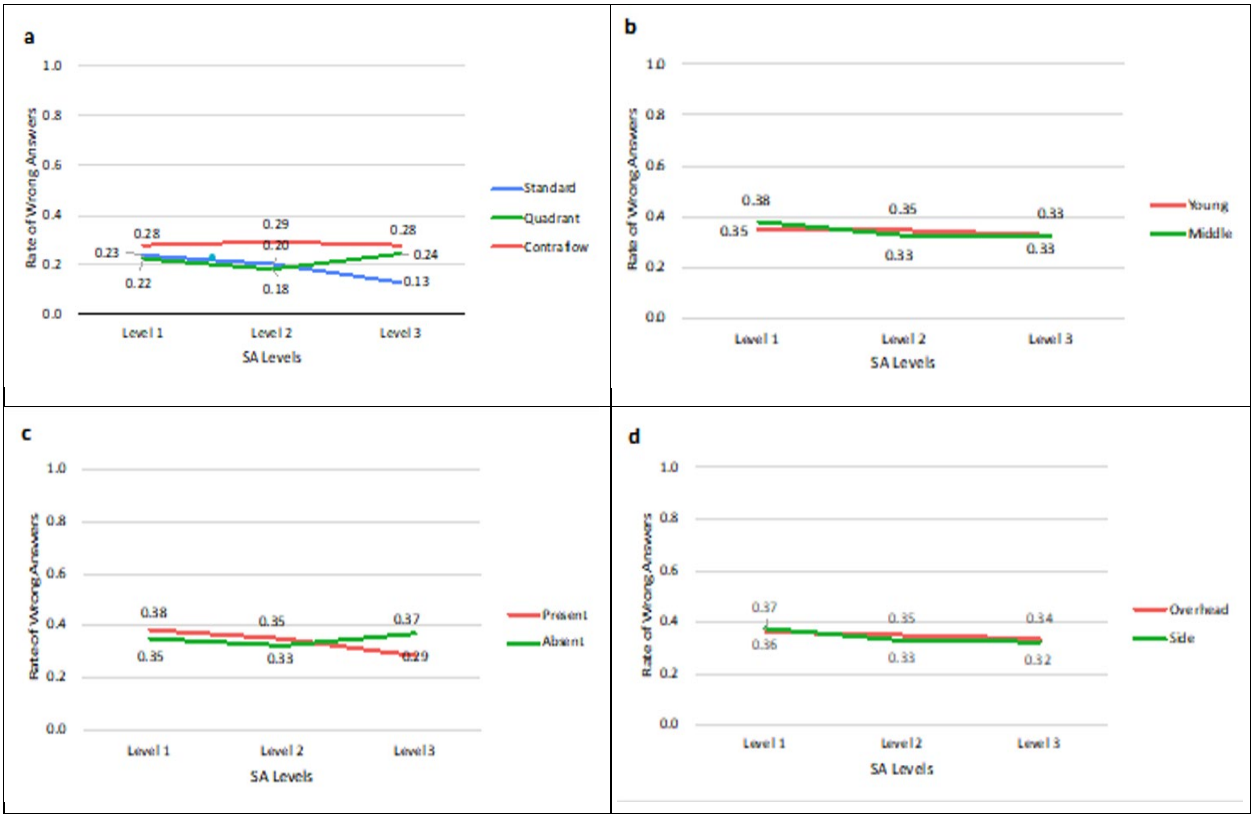


Figure 8: (a) Percent wrong answers for different scenarios; (b) precent wrong answers for different age groups; (c) percent wrong answers for different lane assignment sign use; (d) percent wrong answers for different decision point sign positions.

3.3. Driver Workload

The experiment yielded 48 sets of demand rankings and 288 sets of demand ratings across trials. During the experiment, one set of subject demand rankings, and demand ratings at the close of four trials, were missed. These values were replaced with mean values for all other subjects assigned to and tested under the same signage conditions.

Based on *driver workload* measures, the intersection scenarios each differed; however, age group and sign use, and placement appeared to have no major effect.

Figure 9 (a)(b)(c)(d) presents the mean cognitive workload scores for the different scenarios, age groups, and use and placement of signs, accordingly. Consistent with the SA responses, the cognitive workload of drivers appeared to differ among scenarios, including: Standard: 32.77; Quadrant: 36.19; and Contraflow: 38.86. Age group and sign use and placement appeared to have little-to-no-effect on workload. From analysis of the demand component rankings, we observed driver perceptions of cognitive load to be primarily influenced by their own performance and mental demands, but less by physical demands and frustration (see Figure 10).

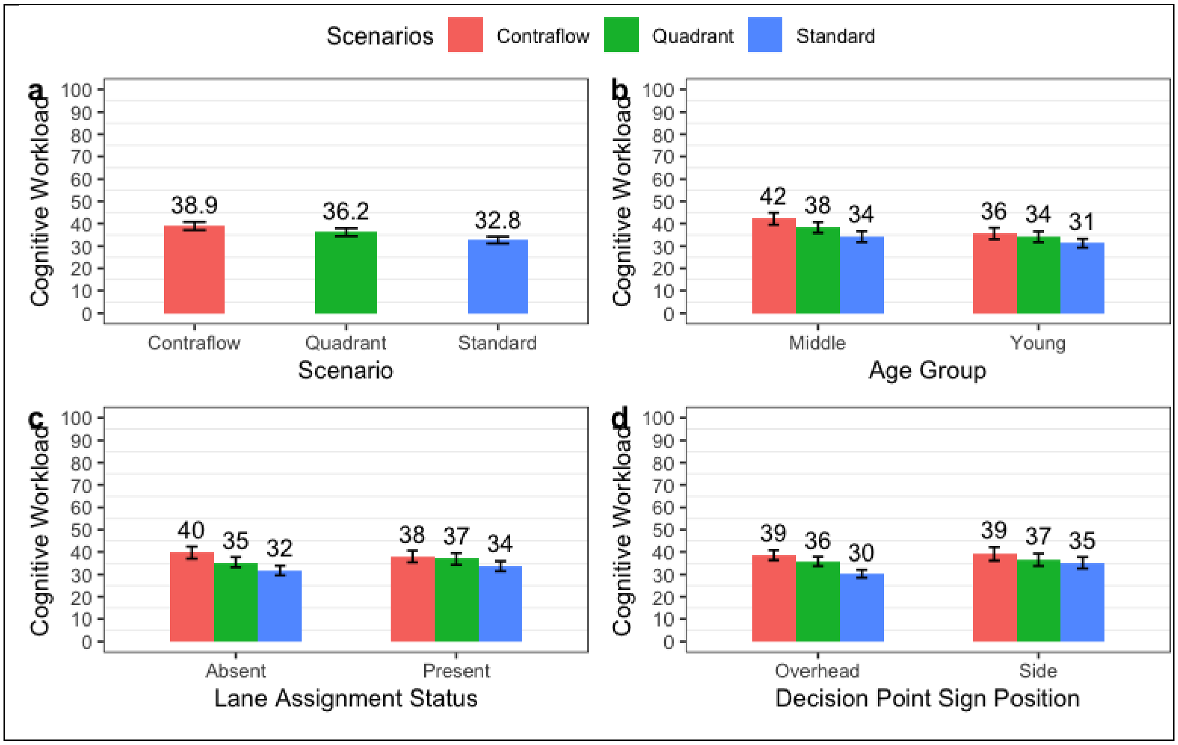


Figure 9: (a) Mean cognitive workload for different scenarios; (b) Mean cognitive workload by scenario type and age group; (c) Mean cognitive workload by scenario type and lane assignment sign use; (d) mean cognitive workload by scenario type and decision point sign position.

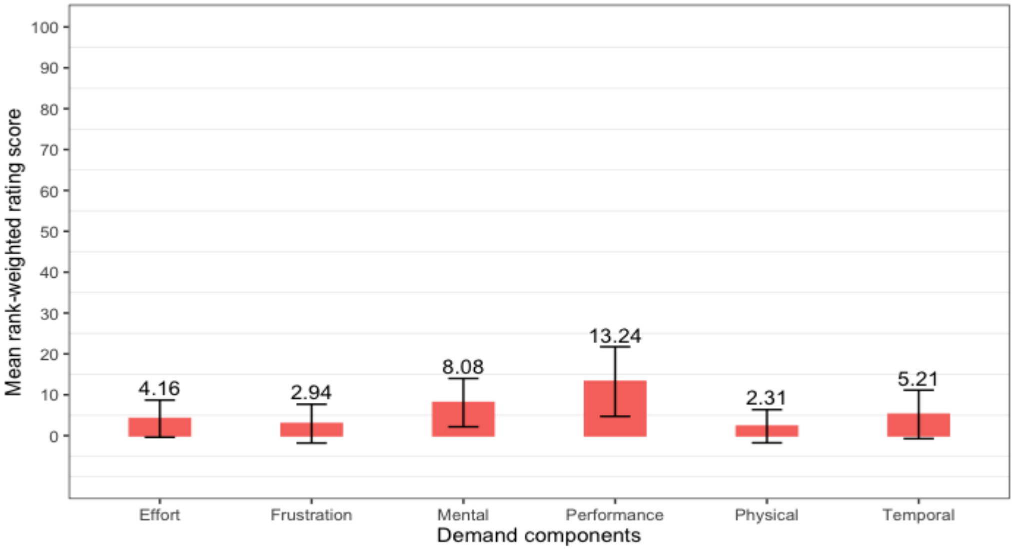


Figure 10: Mean of rank-weighted ratings for six demand components.

Given that the NASA-TLX is a continuous response, and all experiment independent variables were categorical, we applied a multi-way ANOVA to the workload dataset. The diagnosis of TLX scores indicated no normality assumption violation for the parametric test (Shapiro–Wilk test: $p = 0.24$). According to Larson (2008), an ANOVA applied to normal data, even with heterogeneous variance among settings of predictors, is robust for balanced or near-balanced experimental designs. Therefore, the ANOVA test was considered valid for the workload data.

Results were consistent with the SA outcomes, including significant differences in cognitive load under different interchange scenarios ($F_{(2,236)} = 10.73, p = 3.46 \times 10^{-5}$). Post-hoc analysis using Tukey's HSD tests revealed the cognitive workload at Standard intersections to be significantly lower than for the Quadrant ($p = 2.69 \times 10^{-2}$) and Contraflow ($p < 1 \times 10^{-4}$) configurations. However, the cognitive workload of drivers did not differ significantly among the Contraflow and Quadrant conditions ($p = 0.11$). In addition, there were no significant differences in cognitive workload detected among age groups ($F_{(1,41)} = 1.11, p = 0.30$) and sign use ($F_{(1,41)} = 1.75 \times 10^{-2}, p = 0.90$) and placement ($F_{(1,41)} = 0.23, p = 0.63$). Having noted this, when we applied the ANOVA to different demand components, we also observed an almost identical pattern of results as with the TLX scores, except for performance ratings. Consequently, we inferred that any differences in driver cognitive workload are likely the result of the combined effect of various demands. Finally, no interactions were detected among the interchange designs and use and placement of signs for cognitive workload.

3.4. Driver Performance

3.4.1. Absolute Speed Deviation (ASD)

According to Smith (1979), when the visual angle of a stimulus (as subtended at the retina) is 0.007 radians, words can be read clearly. For the signs in the experiment driving scenarios, the driver's virtual vehicle needed to be within 34.84 meters of a side-mounted sign and 46.46 meters of an overhead sign. These values are referred to as foveal viewing distances for the lane assignment and/or decision point signs.

For driver performance assessment, we computed the mean ASD within the foveal distance for the different signage conditions. From all experiment trials, we obtained 288 mean values of ASD within foveal distances without any missing values. The mean ASD for junction sign exposure across participants was 2.66 Km/h with a standard deviation of 3.04; while the mean ASD for lane assignment sign exposure across participants was 3.12 Km/h with a standard deviation of 3.30.

Absolute Speed Deviation (ASD) is the difference in the driver's speed from the posted speed limit. The smaller the difference in speeds, the less likely the subject is expecting a forthcoming movement change.

Figure 11 (a)(b)(c)(d) presents the mean of ASD for different sign exposures, age groups and use and placement of signs, accordingly. Referring to Figure 11 (a), for both junction and lane assignment sign exposures, differences in ASDs among interchange scenarios appear prominent. On average, drivers had the lowest mean ASD in the Contraflow scenario but the highest average ASD in the Quadrant scenario. Figure 11 (b) & (c) reveal that age group did not play an important role in ASD. It was expected that the use of lane assignment signs might also have an impact on ASD; however, Figure 11 (d) suggested that placement of the sign had little effect on the ASD response.

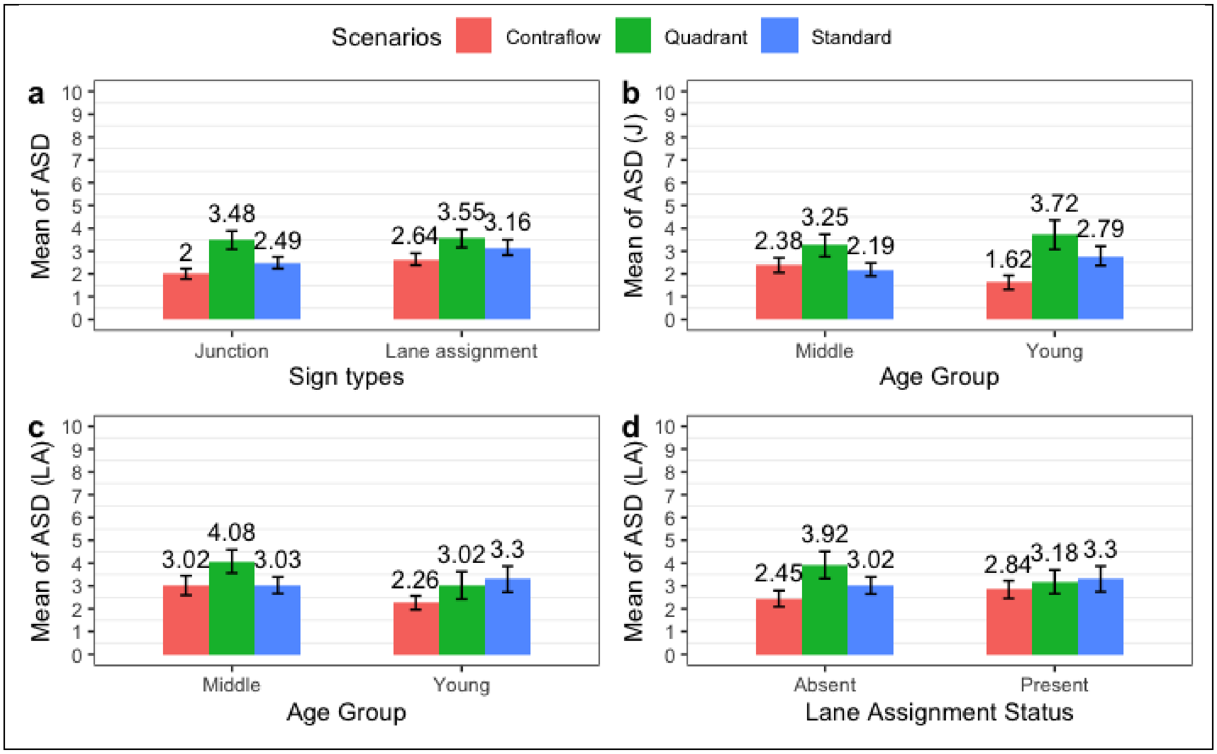


Figure 11: (a) Mean ASD for different signs; (b) mean ASD within junction sign foveal distance by scenario type and age group; (c) mean of ASD within lane assignment sign foveal distance by scenario type and age group; (d) mean of ASD within lane assignment sign foveal distance by scenario type and use of lane assignment sign.

Given that the ASD is a continuous response, we applied a multi-way ANOVA to the response measure, as captured within the different sign foveal distances. ANOVA results revealed driver ASD, when exposed to junction signs, to significantly differ among interchange scenarios ($F_{(2,238)} = 8.55, p = 2.6 \times 10^{-4}$). Post-hoc analyses revealed specific differences among the Contraflow and Quadrant scenarios, with Contraflow GSIs ASD being significantly lower than Quadrant GSIs. This indicates that drivers were not as aware that a likely lane change was imminent at the junction point sign for the Contraflow as compared to the Quadrant. ***This is likely because of the additional lane change information cue on the junction sign for the Quadrant which could also be included on other junction signs for novel intersections to provide better cues to drivers.*** There were no significant differences in driver ASD when they were exposed to lane assignment signs under the various experimental conditions.

Driver ASD was significantly lower at contraflow GSIs. This is likely due to the additional lane assignment information on the junction cueing drivers earlier in the decision process.

3.4.2. Maximum Deceleration (MD)

Figure 12 (a)(b)(c)(d) presents the mean of the MD response for different interchange scenarios, age groups, and use and placement of signs, accordingly. Similar to the other responses we analyzed, for MD, there appeared to be differences among the intersections, including: Standard: 3.47 m/s^2; Quadrant: 3.69 m/s^2;

and Contraflow: $1.38\ m/s^2$. However, graphical analysis indicated little difference in the MD value for other factors, including age group, sign use and placement.

Since MD was also a continuous variable, the ANOVA is a preferred method of analysis with categorical predictors. A likelihood ratio test revealed the variance component for the subject term to be zero, implying no random effect in the model and that other fixed effects accounted for the majority of variation. In this case, our mixed-effect model was reduced to a fixed-effects model.

Maximum Deceleration (MD) is the maximum rate of change in speed measured along the entire route. The larger the deceleration rate, the more uncertainty you would expect during a given trial.

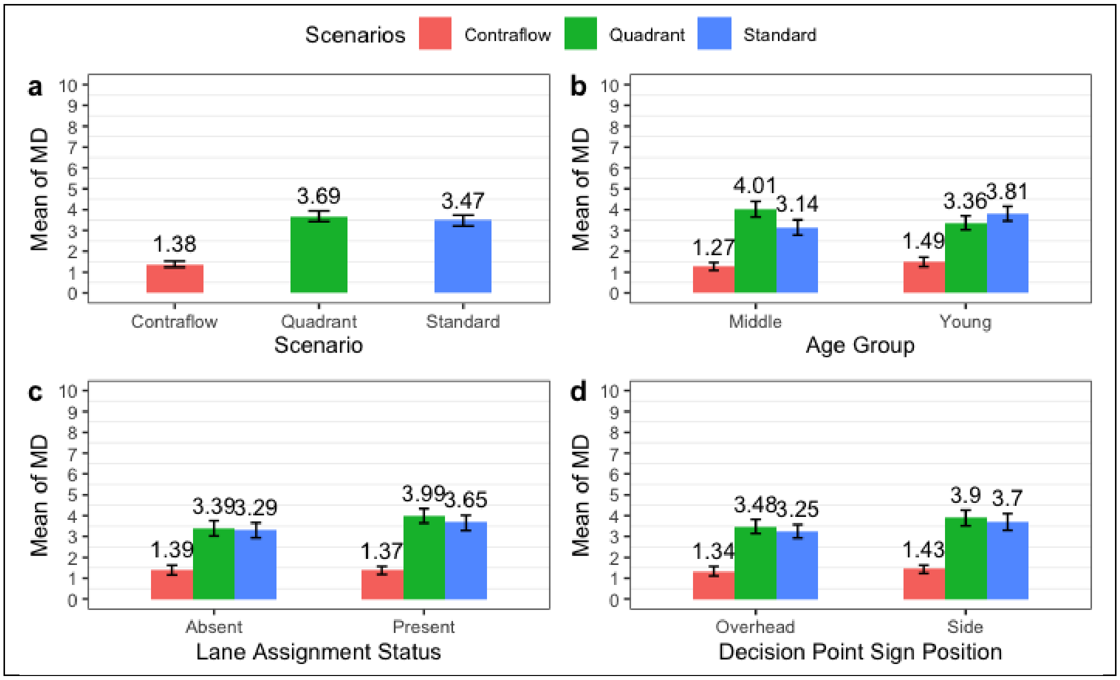


Figure 12: (a) Mean MD for different interchange scenarios; (b) mean MD by scenario type and age group; (c) mean MD by scenario type and lane assignment sign use; (d) mean MD by scenario type and decision point sign position.

ANOVA results revealed a significant difference in the MD of drivers in deceleration sections of the different interchange scenarios ($F_{(2,238)} = 32.33, p = 2.28 \times 10^{-13}$). Post-hoc analyses indicated that this difference stemmed from the fact that the MD for the Contraflow was significantly lower than for both the Standard and Quadrant interchanges. Although the MD rate is higher for the other two intersection forms, the authors do not necessarily believe there is more certainty in navigating the Contraflow. ***The differed is likely due to the deceleration segment of the contraflow beginning far from the point of interchange, while the quadrant and standard intersection designs have deceleration segments located at the periphery of the point of interchange.*** Last, when looking at other factors, there was no statistical evidence of interactions among the different independent factors in the ANOVA model.

The Contraflow yielded significantly lower MD rates compared to the other intersection forms tested; however, this is likely due to the location of the deceleration segment to the point of interchange.

3.4.3. Lane-Changing Position (LCP)

Figure 13 (a)(b)(c) shows the locations at which drivers performed lane changes in three different interchange scenarios. We observed that in the Standard interchange scenario, drivers changed lanes primarily in Segment 1 (at the decision point sign) and Segment 2 (at the lane assignment sign), while only one driver performed a lane change in the deceleration segment. The Contraflow scenario also had the highest number of drivers changing lanes in Segments 1 and 2, but as many as 11 drivers changed lanes during the deceleration segment. Different from the Standard and Contraflow scenarios, the Quadrant produced the greatest number of drivers (70) changing lanes in Segment 3 (at the junction sign), while only one driver made a lane change in the deceleration segment.

Lane-Change Position (LCP) measures where the response to get into the left-most lane takes place relative to the entry of the left turn bay. This provides a measure of when subjects react to the various signs and on-road stimuli.

Figure 13 suggested differences among the three interchange scenarios and the use and placement of signs. Here, it should be noted that the lane assignment sign appeared in Segment 2 for the Standard and Contraflow scenarios, but was presented in Segment 1 for the Quadrant scenario. In general, the descriptive statistics indicated that drivers in the Standard and Contraflow scenarios made lane changes after seeing guidance on the lane assignment and decision point signs. However, in the more complicated Contraflow scenario, many drivers failed to perform a proper lane changing even when they reached the deceleration segment. In the Quadrant scenario, ***the junction sign also displayed some lane information, which may provide insight into why so many drivers changed lanes at that location (in addition to why they may have such different ASD and MD rates).***

Although not significant, based on driver LCP, subjects tended to lane change earlier at the junction point sign for Quadrants vs. lane change signs at the other two GSIs.

The number of drivers performing lane-changes in each road segment was a discrete quantitative variable. Consequently, we applied nonparametric statistical methods to the response measures, specifically the Kruskal-Wallis rank sum test. Results revealed no significant difference in driver lane changing behavior among the different interchange scenarios ($\chi^2 = 0.47, p = 0.79$); however, as observed from the descriptive statistics, there were differences in the number of drivers changing lanes at different road segments ($\chi^2 = 19.97, p = 1.8 \times 10^{-2}$).

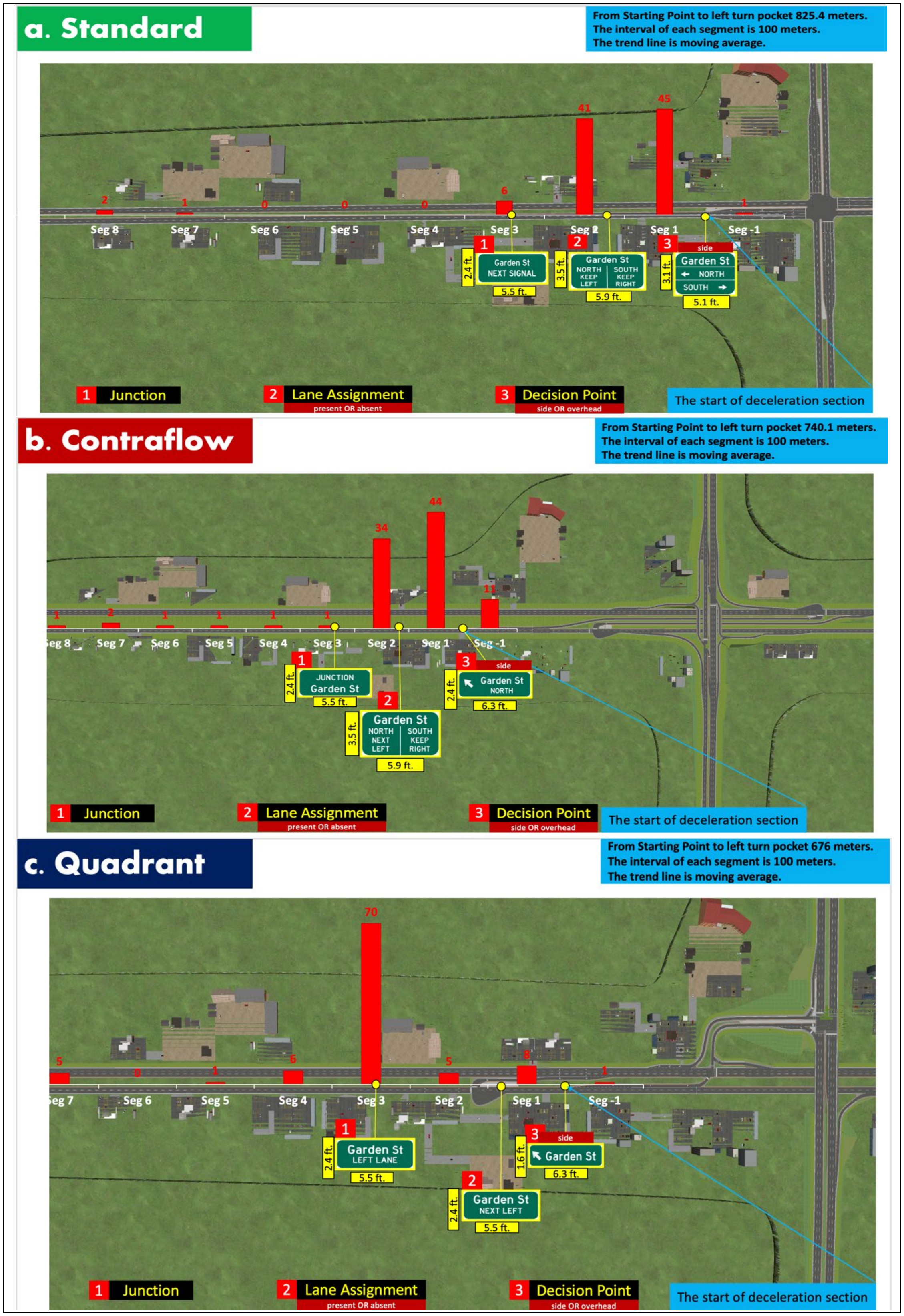


Figure 13: (a) LCP and use and placement of signs in Standard scenario; (b) LCP and use and placement of signs in Contraflow scenario; (c) LCP and use and placement of signs in Quadrant scenario.

3.5. Percentage Change of Pupil Size

For this analysis, we used right-eye pupil size data. As a result of a device connection issue, a large amount of left-eye pupil size data was lost during the experiment. However, it is common practice for PCPS analyses to only make use of observations for one eye. According to Kret & Sjak-Shie (2019), pupil diameter should be between 1.5 and 9 mm. Consequently, we considered any values outside of this range as outliers and the observations were removed from the dataset. On this basis, for foveal viewing of each sign type (junction, lane assignment, and decision point), we obtained 236 valid PCPS data points.

Figure 14 (a)(b) shows the mean PCPS responses during different sign exposures, as well as the use of the lane assignment sign, accordingly. In Figure 14 (a), the mean driver PCPS response had a similar trend for the junction and decision point signs. Regarding the interchange scenarios, on average, driver PCPS was greater in the Quadrant scenario than the Standard and Contraflow scenarios. It is possible that drivers were less familiar with the Quadrant scenario vs. the Standard interchange, leading to elevated visual attention. Here, it is important to note that the decision point signs for the Quadrant scenario were closer to the interchange than in the Contraflow scenario. During lane assignment sign exposure, there appeared to be little if any difference in PCPS among the various interchange scenarios. Figure 14 (b) indicated that there might have been an interaction between the interchange scenarios and lane assignment sign use. In general, the presence of the lane assignment sign in the Contraflow scenario appeared to draw greater driver attention.

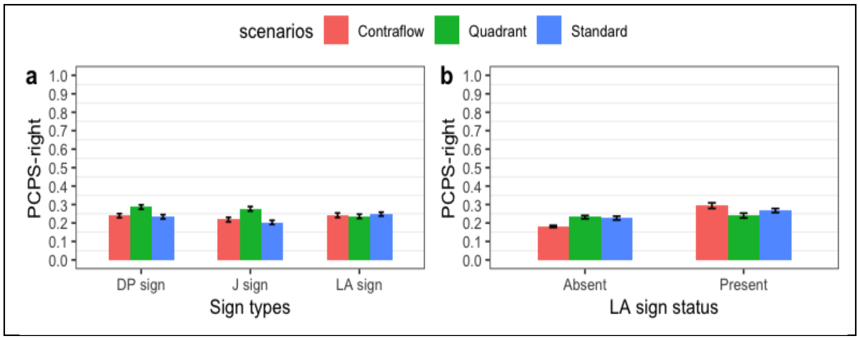


Figure 14: (a) Mean PCPS for foveal viewing of different sign types among interchange scenarios; (b) mean PCPS for foveal viewing of lane assignment signs when in use vs. no viewing.

Figure 15 shows the PCPS response for all types of signs crossed with driver age group. The PCPS for middle-age drivers appeared to be consistently higher than for young drivers. However, there appeared to be limited variation in PCPS responses within the same age group across the different sign types.

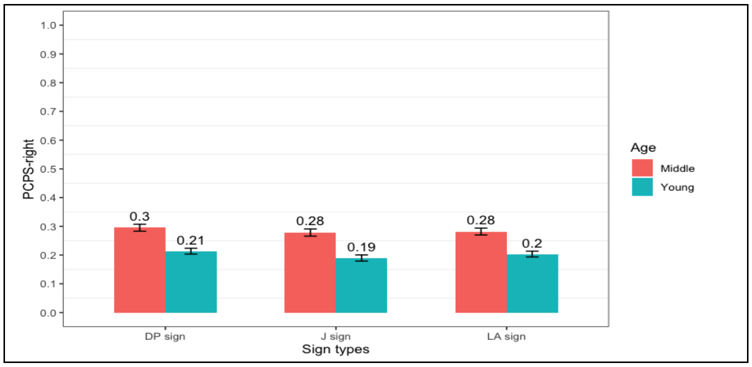


Figure 15: Mean PCPS of different signs by drivers of different ages.

Due to deviations of the response data from parametric test assumptions, a square-root transformation was applied to the PCPS observations (i.e., $Y' = \sqrt{Y}$) and a multi-way ANOVA was conducted for the three sign types. Test results indicated that both age groups ($F_{(1,234)} = 6.05, p = 1.86 \times 10^{-2}$) and interchange scenario ($F_{(2,233)} = 10.28, p = 5.76 \times 10^{-5}$) significantly affected driver PCPS in foveal viewing of junction signs. The PCPS during foveal viewing of lane assignment signs was significantly influenced by age group ($F_{(1,234)} = 4.15, p = 4.91 \times 10^{-2}$) and the interchange scenario by lane assignment sign use interaction ($F_{(2,233)} = 3.10, p = 4.72 \times 10^{-2}$). Only the interchange scenario factor ($F_{(2,233)} = 6.00, p = 2.97 \times 10^{-3}$) had a significant effect on PCPS for foveal viewing of decision point signs.

3.6. Correlation Analysis

In addition to the multi-way ANOVA, we also performed a correlation analysis on the SA and NASA-TLX scores. Given the discrete nature of the SA scores, we computed Kendall's tau correlation coefficients to identify any statistically significant associations based on ranks of the responses. Results revealed no significant correlations between SA and cognitive load (see Table 2; Kendall's tau coefficient of r=-0.055). We also conducted correlation analyses on the SA scores and TLX demand component ratings. No non-parametric correlation coefficients were greater than 0.1 or less than -0.1. These findings indicate that the SA and workload measures are complementary in terms of analysis of human performance and both maybe necessary to elucidate different effects of highway designs on driver behavior and responses.

Based on correlation analysis, SA and workload are complimentary, meaning both are likely necessary to differentiate designs on driver behavior and response.

Table 2: Non-parametric correlation analysis results

Factors	Kendall's tau coefficient
SA Score & NASA-TLX Total Score	-0.055
SA Score & Mental	-0.084
SA Score & Physical	-0.014
SA Score & Temporal	0.016
SA Score & Performance	-0.046
SA Score & Effort	-0.062
SA Score & Frustration	0.043

3.7. Machine Learning Models

Comparison was made of several well-known ML classifiers (SVM, KNN, DT and RF) for predicting driver errors and error-free performance based on underlying cognitive and visual behaviors under the various interchange and signage conditions. Based on classification model training and hyperparameter tuning, we found that several classifiers can achieve high performance with the settings identified in Table 3.

Table 3: Model hyperparameters

Model	Hyperparameters
SVM	kernel=sigmoid, C=1
KNN	neighbors=4
DT	max depth = 3
RF	estimators=20, max depth=3

Table 4 presents the performance for each classification model in training and testing. We found the DT model to produce the highest prediction accuracy, sensitivity/recall, and AUC (aggregate performance) values. Consequently, this analysis and interpretation of model results primarily focuses on the DT approach.

Table 4: Model performance

Model	Training (80%)			Testing (20%)		
	Accuracy	Recall	AUC	Accuracy	Recall	AUC
SVM	0.68	0.23	0.54	0.66	0.26	0.59
KNN	0.70	0.16	0.53	0.67	0.17	0.59
DT	0.77	0.61	0.72	0.83	0.61	0.79
RF	0.72	0.11	0.53	0.67	0.17	0.59

Figure 16 presents the DT structure developed based on our test dataset (58 samples). For this analysis, the maximum depth of the tree was set to three layers of branching preceding classification of all data records (to avoid model overfitting to training data). The classes of driver behavior outcomes were coded using two colors with “orange” representing error-free driving and “blue” for erroneous driving. The greater the hue saturation at the nodes in the tree and leaves (at the base layer), the greater the records that can be accurately classified. The nodes with “white” color represent the highest degree of entropy (disorder) in record classification.

To facilitate the DT construction, the categorical experiment variables were encoded as integers. Figure 15 reveals three variables to be primary predictors of driver behavior in the DT construction, including: (1) the interchange scenario (0-Standard, 1-Contraflow, 2-Quadrant); (2) the NASA-TLX score; and (3) the lane assignment sign use (0-present, 1-absent). The Gini index represents the degree of information provided by each factor at a specific node in the tree with the index ranging from 0 (maximum information) to 0.5 (minimum information). A Gini index value of 0 (see Figure 16, Node ①) indicates homogenous classification of records; otherwise, the classification is non-homogeneous for all other index values. When the Gini index was 0.5 (Figure 16, Node ④), there is no information provided by the node for further data records entering the node. At this node, the error and error-free classes had equal numbers of observations.

The DT analysis process also supports calculation of an importance score for each factor that enters the tree for classification of driver behavior observations. Based on Géron (2019), the importance of a factor is computed as the (normalized) total reduction of entropy (classification uncertainty) provided by that factor, which is also known as the Gini importance. Figure 17 presents the importance score for each factor in the DT, and Table 5 shows the confusion matrix with a test dataset with one false positive and nine false negatives. It can be observed that the interchange scenario type was most important in predicting erroneous driving outcomes, while the NASA-TLX and lane assignment sign use also played important roles in classification.

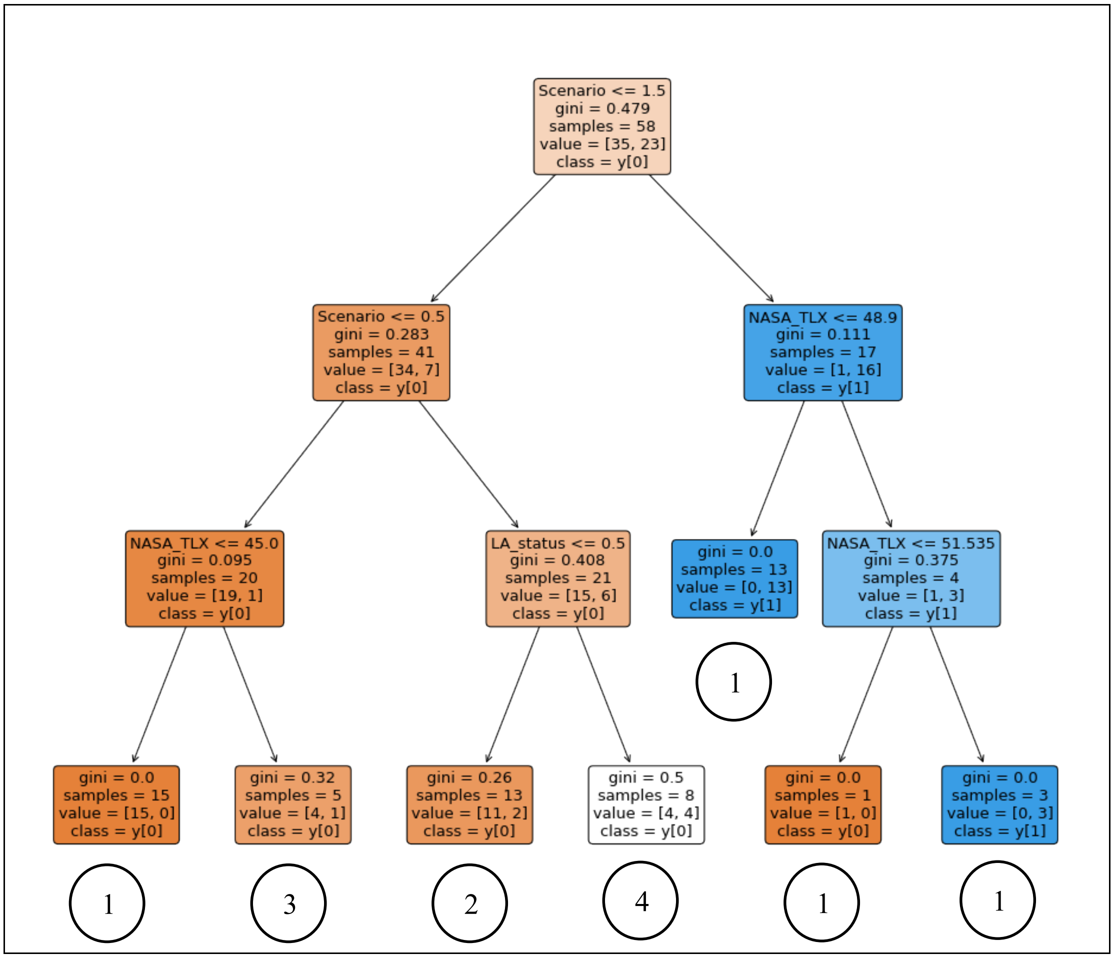


Figure 16: Structure of DT built on test dataset.

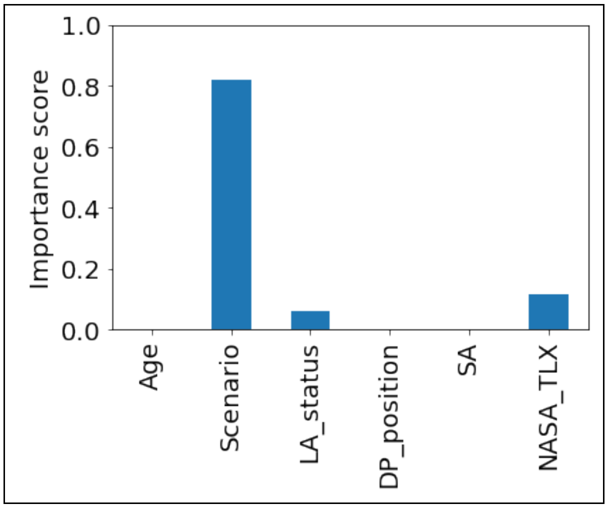


Figure 17: Experimental factor (x-axis) importance scores (y-axis). Higher scores represent greater predictive utility of factors among all factors entering DT model.

Table 5: Common Confusion matrix (DF test dataset)

		Actual class	
		Error	Error-free
Predicted class	Error	TP = 14	FP = 1
	Error-free	FN = 9	TN = 34

4. Conclusions

4.1. Response to Research Objectives

Considering the limitations of traditional interchange designs for managing levels of traffic congestion, alternative interchange designs are becoming more and more prevalent. GSIs are one subgroup of alternative interchange design that are being investigated for the impact of novel geometries for congestion reduction. However, some research has shown that certain GSIs have the potential to increase the incidence of WWD compared to traditional standard intersections. The use of signage can be an effective countermeasure for improving driver awareness of roadway configurations and thereby reduce potential incidents. Consequently, the objectives of this study were to assess the effects of select GSI designs compared to standard intersection designs in terms of driver situation awareness, cognitive workload, vehicle control performance, and visual attention allocation under various routing sign conditions, including junction, lane assignment and decision point signs. We also proposed to model driving outcomes (error-free or WWD), based on driver behaviors exhibited in responding to the interchange configuration and signage conditions.

Overall, the results of this study partially supported Hypotheses 1 and 2. The use and placement of signs at the simulated GSIs did not result in significant differences in driver subjective responses such as SA and cognitive workload. However, the objective measures of driver performance and visual behavior (PCPS) were sensitive to the signage manipulations. To be more specific, use of lane assignment signs at interchanges did not appear to significantly increase driver SA nor decrease driver cognitive workload. Furthermore, overhead mounted decision-point signs did not significantly increase driver SA or decrease cognitive workload, as compared to side-mounted signs. However, driver performance measures revealed that for the standard intersection and contraflow GSI, drivers primarily changed lanes after lane assignment sign use (Segment 2) or decision point sign (Segment 1) presentation. For the quadrant GSI, drivers primarily executed lane changes at the junction sign location (Segment 3) as quadrant junction signs also displayed lane information. Furthermore, analysis of the eye-tracking measures revealed PCPS to significantly increase when lane assignment signs were present and, consequently, this manipulation was more important for the contraflow design. The lack of significance of the decision point sign on driver behaviors could have been due to the overhead mount not being as visually accessible as expected, relative to the side-mounted signs.

Results partially supported Hypothesis 3. The contraflow design led to significantly degraded driver SA, likely due to a lack of driver familiarity with the configuration, as compared to the standard and quadrant interchanges. However, the quadrant design did not differ from the standard intersection in terms of SA. Results on cognitive workload revealed significant increases for drivers at both the contraflow and quadrant interchanges, as compared to the standard intersection. However, there was no difference between the contraflow and quadrant designs in terms of workload. Once again, these findings are likely due to the novelty of the GSI interchanges, lack of driver familiarity, and perceived complexity of navigation of the interchanges. In addition, driver performance responses showed the lowest MD for the contraflow design but the highest MD for the quadrant. The unique design of each scenario could have led to these results. Specifically, the deceleration segment of the contraflow begins far from the point of interchange, while the quadrant and standard intersection designs have deceleration segments located at the periphery of the point of interchange. Related to this, the contraflow design resulted in many more drivers executing a LCP maneuver at or after the opening of the deceleration segment. We also found this interchange design to be highly correlated with the incidence of WWD through the machine learning (DT) model.

On the basis of these results, it can be inferred that the further upstream left in the contraflow design is generally more complex/confusing for drivers than the quadrant GSI or standard intersection. We also suspect that the location of the decision point sign far away from the point of interchange may account for the observed low driver SA, high workload and elevated incidence of driving errors for the contraflow. While not directly tested, it appears providing additional contextual information on the junction sign results in earlier lane change position for the quadrant compared to the contraflow design as well as earlier increases in ASD for the quadrant indicating earlier preparation for a non-traditional movement.

Beyond the roadway configuration manipulations, there was no significant evidence of differences in driver SA, cognitive workload and performance due to age (younger drivers: 18 to 24 years; middle-aged drivers: 25 to 64 years). This result was likely due to a limited age gap between our study groups. From the convenience sample, we observed that 68.75% of subjects were between 18 and 26 years of age. For future study, there is a need to collect additional data on elderly drivers to more conclusively determine whether age has an influence on driver SA responses for different types of interchanges (standard vs. GSIs).

Based on these observations, it is recommended that signing engineers develop novel junction sign configurations, or provide additional guidance signs upstream of the decision point, for intersections with non-traditional movements in order to offset low driver SA and high cognitive workload, and to support timely lane changing behavior. The quadrant GSI design appears to be a feasible alternative to standard intersections with or without lane assignment signs and when using side-mounted decision point signs and providing lane information on the junction sign. Consequently, the results of this study provide some guidance for highway systems engineers on the need for novel signage designs to ensure effective driver information processing under unique highway configurations with performance comparable to standard intersections. It is inferred that driver performance compared to standard intersections is similar at intersection forms with a non-intuitive turning movement (e.g. turn left to go right), whereas drivers at intersection forms which require advanced lane changing (e.g. contraflow and displaced left turns) may require additional guidance beyond that provided at standard intersections.

Last, in the experiment design, the research team noted that the contraflow design findings should have direct application to the continuous flow intersection (not studied in this effort) because they both require a left turn upstream of the normal intersection point. Therefore, similar to the contraflow, the findings from this effort support the use of additional lane information on signs at the junction point to help aid drivers in reduced lane changing prior to the decision point.

4.2. Study Limitations and Future Research

Based on the available experiment resources, we elected to expose subject drivers to two specific types of GSIs and standard intersections. Although contraflow and quadrant designs are among the more frequently applied GSIs in practical application, other geometries (e.g., trumpet interchange and diamond interchange) have yet to be investigated for SA and cognitive workload impacts on drivers leading to WWD incidents. Therefore, all observations in this study are restricted to the specific test GSIs or intersection forms with similar features (i.e. turn left to go right and advanced lane changing). Further experiments are needed to generate more general conclusions across additional types of GSIs.

Similarly, lane assignment and decision point signs are only two forms of transportation guidance signs for which we explored use and placement, respectively. Further investigation of use and placement of other types of guidance signs in different GSI designs is worthwhile to draw more general conclusions about the utility of signage for managing driver behaviors in interchange negotiation.

Regarding the analysis of driver age for influence on behavior and performance with the various GSI designs and standard intersection, the lack of a senior driver group (65+ years of age) as part of the experiment sample might have attenuated any differences in SA, cognitive load, and vehicle control with age. Had we been able to recruit participants with a broader age range, differences in these responses might have been pronounced for the interchange geometries and the signage configurations. Although we proposed to examine senior driver behavior and performance, this was not possible due to the COVID-19 pandemic conditions that occurred during the course of this project. In addition to driver age, we note that the particular sample produced slightly high PCPS values (relative to the prior literature). This observation is likely not attributable to individual subject characteristics. Rather, the fact that the experimental environment allowed for limited lighting control, and inconsistencies occurred in illumination intensity at different time periods during the experiment, might have influenced PCPS measures.

The machine learning modeling approach taken in this study appears to have utility for identifying roadway and signage factors that are most influential in driver behaviors and, consequently, predictive of driver errors. Unfortunately, the study only yielded 288 data records for model training and testing, which is a comparatively small sample size for any machine learning analysis. For effective training of machine learning classifiers and demonstration of validity, a much larger data set is needed to generate more reliable results.

5. References

Abdelrahman, A. (2020). Safety Evaluation of Innovative Intersection Designs : Diverging Diamond Interchanges and Displaced Left-turn Intersections.

Atiquzzaman, M., & Zhou, H. (2018). Modeling the Risk of Wrong-Way Driving Entry at the Exit Ramp Terminals of Full Diamond Interchanges. *Transportation Research Record*, 2672(17), 35–47.

Baratian-Ghorghi, F., Zhou, H., Jalayer, M., & Pour-Rouholamin, M. (2015). Prediction of Potential Wrong-Way Entries at Exit Ramps of Signalized Partial Cloverleaf Interchanges. *Traffic Injury Prevention*, 16(6), 599–604. https://doi.org/10.1080/15389588.2014.981651

Braam, A.C. (2006). Wrong-way Crashes: Statewide Study of Wrong-way Crashes on Freeways in North Carolina. North Carolina Department of Transportation, Raleigh.

Brown, H., Edara, P., Hartman, G., & Chlewick, G. (2020). Alternative Intersection Design and Selection (No. Project 20-05, Topic 50-11).

Carter, D., & Wang, I. (2007). Evaluating options to increase specific service (logo) signs from six businesses to nine businesses per service. Final Report, University of North Carolina Highway Safety Research Center, Chapel Hill, NC.

Chase, R.T., Cunningham, C., Vaughan, C., Warchol, S., & Lee, T. (2020). Reasonable Alternatives for Grade-Separated Intersections. Report No. FHWA/NC/2018-20, North Carolina Department of Transportation, Raleigh.

Chilukuri, V., S. Smith, M. Trueblood, and T. Ryan, Diverging Diamond Interchange Performance Evaluation, Report No. OR11-012, Missouri Department of Transportation, Jefferson City, 2011.

Claros, B., Edara, P., Sun, C., Brown, H. (2015). Safety Evaluation of Diverging Diamond Interchanges in Missouri. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2486. Transportation Research Board of the National Academies, Washington, D.C, pp. 1–10.

Claros, B., Edara, P., & Sun, C. (2017). When driving on the left side is safe: Safety of the diverging diamond interchange ramp terminals. *Accident Analysis and Prevention*, 100, 133–142. https://doi.org/10.1016/j.aap.2017.01.014

Copelan, J.E. (1989). Prevention of Wrong-way Accidents on Freeways. California Department of Transportation, Sacramento. FHWA/CA-TE-89-2. http://www.ce.siue.edu/faculty/hzhou/ww/PREVENTION-OF-WRONGWAY-ACCIDENTS-ON-FREEWAYS.pdf

Cottrell Jr, B. H., & Edara, P. (2011). Evaluation of adding distance information to freeway-specific service (logo) signs. *Journal of transportation engineering*, *137*(11), 782-788.

Cunningham, C., S. Warchol, and B. Schroeder, Exploring Corridor Operations in the Vicinity of a Diverging Diamond Interchange (DDI), Report No. FHWA/NC/2014-13, North Carolina Department of Transportation, Raleigh, 2016b.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462-492.

Eyler, D. R. (2005). Arterial interchange. *Transportation Research Record*, 1912, 65–71. https://doi.org/10.3141/1912-08

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis and Prevention*, 37(3), 461–472. https://doi.org/10.1016/j.aap.2004.11.003

FHWA. (2009). Manual on uniform traffic control devices. CHAPTER 2E . GUIDE SIGNS — FREEWAYS AND EXPRESSWAYS Section 2E . 01 Scope of Freeway and Expressway Guide Sign Standards Standard : Section 2E . 02 Freeway and Expressway Signing Principles Section 2E . 04 Color of Guide Signs Standard

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. " O'Reilly Media, Inc.".

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics (Vol. 1). New York: Wiley.

Halim, Z., & Rehan, M. (2020). On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning. *Information Fusion*, *53*, 66-79.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*. https://doi.org/10.1016/S0166-4115(08)62386-9

Hughes, W., Jagannathan, R., Sengupta, D., & Hummer, J. (2010). *Alternative intersections/interchanges: informational report (AIIR)* (No. FHWA-HRT-09-060). United States. Federal Highway Administration. Office of Research, Development, and Technology.

Inman, V. W., Katz, B. J., & Hanscom, F. R. (2006). Navigation signing for roundabouts. *Transportation Research Record*, 1973, 18–26. https://doi.org/10.3141/1973-05

Jabon, M., Bailenson, J., Pontikakis, E., Takayama, L., & Nass, C. (2010). Facial expression analysis for predicting unsafe driving behavior. *IEEE Pervasive Computing*, *10*(4), 84-95.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. In *Theoretical Issues in Ergonomics Science* (Vol. 5, Issue 2). https://doi.org/10.1080/1463922021000054335

Kaber, D. B., Liang, Y., Zhang, Y., Rogers, M. L., & Gangakhedkar, S. (2012). Driver performance effects of simultaneous visual and cognitive distraction and adaptation behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(5), 491–501. https://doi.org/10.1016/j.trf.2012.05.004

Kaber, D., Pankok, C., Corbett, B., Ma, W., Hummer, J., & Rasdorf, W. (2015). Driver behavior in use of guide and logo signs under distraction and complex roadway conditions. *Applied Ergonomics*, 47, 99–106. https://doi.org/10.1016/j.apergo.2014.09.005

Leisch, Joel P. (1993). "Grade-Separated Intersections." *Transportation Research Record* 1385.

Lloyd, H. (2016). A Comprehensive Safety Analysis of Diverging Diamond Interchanges.

Lochrane, T., Bared, J., & Zhang, W. (2011). Capacity analysis for planning of junctions (CAP-X). *FHWA, US Department of Transportation*, *2*.

Moler, S. (2002). Stop. You are going the wrong way! *Public Roads* 66 (2), 24e29.

Morena, D.A., & Leix, T.J. (2012). Where these drivers went wrong. *Public Roads*, 75(6):33–41. https://www.fhwa.dot.gov/publications/publicroads/12mayjune/05.cfm

Neuman, T., Nitzel, J., Antonucci, N., Nevill, S., & Stein, W.(2008). A Guide for Reducing Head-On Crashes on Freeways. NCHRP Report 500, Vol 20. https://doi.org/10.17226/23088

North Texas Tollway Authority (NTTA). (2009). Keeping NTTA Roadways Safe: Wrong-way Driver Task Force Staff Analysis. NTTA, Plano.

Osman, O. A., Hajij, M., Karbalaieali, S., & Ishak, S. (2019). A hierarchical machine learning classification approach for secondary task identification from observed driving behavior data. *Accident Analysis & Prevention*, *123*, 274-281.

Qiao, F., Liu, X., & Yu, L. (2007). Using Driving Simulator for Advance Placement of Guide Sign Design for Exits along Highways. Department of Transportation Studies Texas Southern University, January.

Richard, C. M., & Lichty, M. G. (2013). Driver expectations when navigating complex interchanges (No. FHWA-HRT-13-048). Turner-Fairbank Highway Research Center.

Salamati, K., Schroeder, B. J., Rouphail, N. M., Cunningham, C. M., Zhang, Y. & Kaber, D. B. (2012). Simulator study of driver responses to pedestrian treatments at multilane roundabouts. *Transportation Research Record (Journal of the Transportation Research Board)*, Vol. 2312, 67-75.

Shin, E.K., J. H. Lee, J. H. Kim, J. S. Kim, and Y. W. Jeong. Two-Level Signalized Intersection. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2060, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 53–64

Smith, S. L. (1979). Letter size and legibility. *Human factors*, *21*(6), 661-670.

Song, Y., Liu, H., & Ge, S. (2014). Quantitative evaluation of the position of an interchange advance guide sign based on traffic simulation technology. In *CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems* (pp. 2653-2659).

Stanić, Branimir, Vladan Tubić, and Nikola Čelar. "Design and evaluation of a grade-separated intersection: a case study of the proposed Belgrade 'Hipodrom'." *Transportation planning and technology* 34.6 (2011): 625-636.

Stanton, N. A., Hedge, A., Brookhuis, K., Salas, E., & Hendrick, H. W. (2004). Situation awareness measurement and the situation awareness global assessment technique. In *Handbook of human factors and ergonomics methods* (pp. 445-453). CRC Press.

Topolšek, D. (2007). Prevention of wrong-way driving on freeways. *Promet-Traffic&Transportation*, *19*(5), 311-321.

Upchurch, J., Fisher, D. L., & Waraich, B. (2005). Guide signing for two-lane exits with an option lane: Evaluation of human factors. *Transportation Research Record*, 1918, 35–45. https://doi.org/10.3141/1918-05

Vaughan, C., Jagadish, C., Bharadwaj, S., Cunningham, C. M., Schroeder, B. J., Hummer, J. E., Findley, D., & Rouphail, N. M. (2015). Long-term monitoring of wrong-way maneuvers at diverging diamond interchanges. *Transportation Research Record*, 2484(1), 129–139. https://doi.org/10.3141/2484-14

Vicedo, P. (2006). Prevention and management of ghost drivers incidents on motorways: the French experience the contribution of ITS to immediate detection and optimum management of ghost drivers incidents. In: *The 13th ITS World Congress*, London, 2006.

Wang, J. (2018). Relationship between Geometric Elements and Wrong-Way Crashes at Partial Cloverleaf Interchange Terminals. Institute of Transportation Engineers. *ITE Journal*, 88(4), 45-49.

Wortman, R. H., & Fox, T. C. (1986). A reassessment of the traffic signal change interval. University of Arizona.

Yeo, M. V., Li, X., Shen, K., & Wilder-Smith, E. P. (2009). Can SVM be used for automatic EEG detection of drowsiness during car driving?. *Safety Science*, *47*(1), 115-124.

Zahabi, M., & Kaber, D. (2018). Effect of police mobile computer terminal interface design on officer driving distraction. *Applied ergonomics*, *67*, 26-38.

Zahabi, M., Machado, P., Lau, M. Y., Deng, Y., Pankok, C., Hummer, J., Rasdorf, W., & Kaber, D. B. (2017). Driver performance and attention allocation in use of logo signs on freeway exit ramps. *Applied Ergonomics*, 65, 70–80. https://doi.org/10.1016/j.apergo.2017.06.001

Zahabi, M., Nasr, V., Mohammed Abdul Razak, A., Patranella, B., McCanless, L., & Maredia, A. (2021). Effect of secondary tasks on police officer cognitive workload and performance under normal and pursuit driving situations. *Human factors*, 00187208211010956.

Zahabi, M., White, M. M., Zhang, W., Winslow, A. T., Zhang, F., Huang, H., & Kaber, D. B. (2019). Application of cognitive task performance modeling for assessing usability of transradial prostheses. *IEEE Transactions on Human-Machine Systems*, *49*(4), 381-387.

Zhang, Y., Kaber, D. B., Rogers, M., Liang, Y., & Gangakhedkar, S. (2014). The effects of visual and cognitive distractions on operational and tactical driving behaviors. *Human Factors*, 56(3), 592–604. https://doi.org/10.1177/0018720813495279

Zhang, Y., Harris, E., Rogers, M., Kaber, D., Hummer, J., Rasdorf, W., & Hu, J. (2013). Driver distraction and performance effects of highway logo sign design. *Applied Ergonomics*, 44(3), 472–479. https://doi.org/10.1016/j.apergo.2012.10.009

Zhou, H., Zhao, J., Fries, R., Reisi Gahrooei, M., Wang, L., Vaughn, B., Bahaaldin, K., Ayyalasomayajula, B., 2012a. Investigation of Contributing Factors Regarding Wrong-Way Driving Crashes on Illinois Freeways. Publication FHWA-ICT-12-010. FHWA. U.S. Department of Transportation.

Zhou, H., Wang, L., Gahrooei, M.R., Jalayer, M., 2012b. Contributing factors to wrong- way driving crashes in Illinois. ITE Midwestern District Conference and TRB 4th Urban Street Symposium. pp. 24–27.

Zwahlen, H. T., Russ, A., Roth, J. M., & Schnell, T. (2003). Effectiveness of ground-mounted diagrammatic advance guide signs for freeway entrance ramps. *Transportation Research Record*, 1843, 70–80. https://doi.org/10.3141/1843-09